

THIS WEEK

EDITORIALS

SUSTAINABILITY Economic and environmental causes share a common goal **p.288**

WORLD VIEW Hey science education advocates, leave those kids alone **p.289**



GROUND FORCE Lemurs cope with winter by going underground **p.291**

Privacy in the digital age

The proposed European Data Protection Regulation will rightly preserve people's privacy — but, without exceptions for scientific research, it could hinder or prevent medical discoveries.

Some commercial efforts to mine and exploit data come across as creepy. The US retailer Target got so good at identifying expectant parents that it started to post them coupons for deals on baby clothes. Many people dislike Google's practice of targeting advertisements based on analysis of private e-mails and web searches.

The European Parliament is discussing data-privacy legislation that would limit such efforts by curtailing customer profiling and providing a 'right to be forgotten'. But in its well-intentioned zeal, the European Union (EU) law could slow and even prevent the discovery of life-saving medical interventions.

The EU's draft Data Protection Regulation is a much-needed update to the Data Protection Directive, a law passed in 1995 — when 'Amazon' was best known as the name of a river. The first draft of the rewrite contained exemptions for personal data used by scientists, as long as identifiable data were used only when necessary and with measures to protect privacy and limit access. Such protections are already common in research using patient data, and are an important component of biobanks, genetic databases and other scientific resources established for the common good.

The latest news from Brussels has European scientists and funders understandably on edge. The European Parliament's Committee for Civil Liberties, Justice and Home Affairs (LIBE) has provisionally removed many of the exemptions for research, essentially treating scientific institutions in the same way as e-mail spammers.

Under the proposed amendments, a person's identifiable health data could be used only with "specific, informed and explicit consent". It sounds reasonable that people should be asked before researchers use their information. But the wording of the amendment is vastly out of step with how people consent to research and how science is done.

Increasingly, volunteers make their DNA, tissue samples and health information available under 'broad consent'. That means that researchers do not need to get permission for every single use of a patient's records — for instance, to validate a newly discovered genetic biomarker for cancer risk. Biobanks typically obtain broad consent from donors, and the proposed amendments could make it nearly impossible for them to operate.

It is not always possible to get a person's permission to use their personal data for important health research. High numbers of participants can make such an exercise impractical: for example, a 2005 study linking preterm birth with high blood pressure later in life used medical data from more than 300,000 Swedish men involved in a cohort study (S. Johansson *et al. Circulation* **112**, 3430–3436; 2005). Such unconsented research is done under the strictest ethical safeguards, yet the draft law would create an even higher hurdle — and much research would founder as a result.

Most under threat is the next generation of biomedical research, in which scientists hope to make discoveries by mining medical records from health systems such as the UK National Health Service.

Last month, the UK Medical Research Council opened four health-informatics centres to lay the groundwork for such efforts.

The new law will offer some opportunities to researchers. It will be the same across all states in the EU, offering the potential to ease international collaborations by streamlining complex data-protection rules. And the requirement that data holders provide a means for people to access and withdraw their personal data offers researchers a chance to engage study participants through online portals. In

"The wording of the amendment is vastly out of step with how people consent to research."

addition to allowing individuals to drop out with the click of a button, these portals could keep participants abreast of research findings and allow both participants and researchers to request new information. Earlier this year, the Netherlands Twin Register created the MyBiobank mobile app, which has been well received by participants.

The Data Protection Regulation is still taking shape, and the European Parliament will consider more than 3,000 amendments when the LIBE committee convenes. Whatever law the parliament passes must be reconciled with a version taking shape in the Council of the European Union, which is made up of national ministers. There is ample time for the legislation to evolve to be more friendly to research, before it is finalized in 2014. *Nature's* readers will be affected by this law and they still have the chance to influence how it is written. The UK biomedical charity the Wellcome Trust is encouraging scientists in Europe to draw the attention of their local Member of the European Parliament to its statement of concerns (go. nature.com/meb4pm). Join them, and do a little targeted advertising of your own. ■

Science in schools

The US National Center for Science Education teaches researchers how to fight for their cause.

Critics of mainstream science frequently dispute evolution or climate change. Whatever their target, a common tactic is to challenge how well mainstream scientists accept these ideas.

When the anti-evolution Discovery Institute in Seattle, Washington, began a project in this vein, creating lists of scientists who doubt evolution, the pro-evolution National Center for Science Education (NCSE) responded in kind. It collected responses from PhD-level scientists who agreed that there is "no serious scientific doubt" that evolution occurred — but only those who were called Steve or a variant. This light-hearted

list of Stephens, Stephanies and similars now dwarfs the list of doubters, making a clear statement about where mainstream science stands.

That statement does not, and is not intended to, inform scientists. But it buttresses their long-term futures. To ensure that the supply of competent young researchers and policy-makers does not fail, the public should be educated in a vital, unifying principle of biology. Yet teachers are often pressured to keep evolution out of the classroom or to teach it as a scientifically controversial theory, particularly in the United States.

The NCSE, which is based in Oakland, California, is committed to tackling such attacks. It is perhaps most famous for organizing plaintiffs in the 2005 case *Kitzmiller v. Dover*, in which parents in Pennsylvania sued a school board for requiring that intelligent design be taught in public schools. The case was decided in favour of the parents, a ruling that is credited with keeping intelligent design out of classrooms across the United States. But the NCSE has probably had a similar impact in its quieter battles: it provides resources for science advocates, so that they do not have to reinvent the wheel when helping teachers who are told to skip evolution, to misrepresent it as controversial or to teach it alongside unscientific ideas. And the centre adapts to current needs: last year, it branched out to include climate change in its education efforts.

Eugenie Scott, executive director of the NCSE for the past 26 years, announced her retirement last week. The scientific community has much to learn from her example in the fight against pseudoscience. Too often, scientists are ignorant of how students outside their own labs are being educated. In the worst cases, scientists can actually hurt the cause for science education by alienating the people whom they hope to persuade: in their attempts to engage, they may seem condescending or use arcane arguments that fail to connect with teachers, parents, students and other community members.

Science is necessary to defuse anti-science efforts, but not sufficient. Rather than simply deploying artilleries of scientific facts, the NCSE addresses the motivations and tactics of those who would misrepresent research. These individuals and groups are driven not by facts, says Scott,

but by ideologies and identities such as 'fundamentalist Christian' or 'political conservative'. Scott's strategy is to attack what she calls dichotomous thinking: false assumptions that a churchgoer cannot believe in evolution or that a scientist cannot believe in a higher power. When, in 1995, the US National Association of Biology Teachers issued a statement describing evolution as "impersonal" and "unsupervised", Scott and others called successfully for those words to be removed, arguing that science could not address such questions.

"Too often, scientists are ignorant of how students outside their own labs are being educated."

Another strategy is to put together coalitions of people from diverse backgrounds to provide multiple perspectives. Faith-based communities can express concerns about one religious view being favoured over another. Parents can argue for their children's clear thinking and academic futures. Scientists can talk about the scientific process and why accuracy in schools matters, but should also participate, where applicable, as parents, community members or people of faith.

Scientists and their institutions can encourage public outreach. Articulate researchers are no longer frowned on for being able to engage with a broader public, but they are not always supported; their institutions should consider how to recognize those who communicate science to society.

Scientists will need to learn to shift gears. As professionals, they must advocate for their own research, explaining why their grant should be funded or their papers published. When it comes to celebrating science more generally, they should bring the same passion to describing the work that is most likely to engage the public. The late evolutionary biologist Stephen Jay Gould (honoured in the NCSE's list of Steves) established his professional reputation through accounts of his own gritty field work, but popularized science more through his discussions of the work of others. With support from the NCSE and similar efforts, scientists can further not only science education, but science itself. ■

Together we stand

To reach a sustainable future, we must merge economic and environmental agendas.

In the middle years of the last decade, when political concern over climate change was at its height, politicians liked to boast that they had broken the link between economic growth and greenhouse-gas emissions. They hadn't, as the impact of the global financial crisis and the tentative recovery showed: emissions in some places slowed temporarily but have now resumed their worrying rate of increase.

Although countries such as the United Kingdom did manage to reduce their carbon pollution, this was a product of specific policies such as a shift from coal to gas for electricity generation, and not a more fundamental sign of a shift in the economic sands. Protecting the environment is an added cost that many politicians and business leaders would prefer to avoid. Not to bother makes things cheaper. And despite the rhetoric of environmental campaigners, that remains an uncomfortable truth, at least in terms of the climate problem. Carbon emissions are a hallmark of energy use — and it is cheap and available energy that has made the modern world.

The antagonism between protection of profit and protection of the environment will continue for as long as the two are seen as separate pursuits. And although there are signs that people in China are starting to question their nation's pursuit of the first and not the second (see Q. Wang *Nature* 497, 159; 2013), the rapid economic development of countries such as China, India and Brazil raises the stakes and the risk to a sustainable future still further.

Development is a right, of course. Poorer nations are entitled to follow the path to prosperity, especially when their economic growth is measured in terms of reduced childhood mortality and increased access to clean water, as well as industrial output. But it is in all of our interests to find a more sustainable way for them to do so.

Global goals, international targets and multilateral pledges in this arena can sound hollow. There is little evidence, for instance, that the United Nations' Millennium Development Goals have achieved what they set out to do. When these goals expire at the end of 2015, should they be renewed? And, if so, with what? The answer to the first is an unequivocal 'yes'. A goal is something to strive for, as well as a result. Targets set agendas and steer policy. A possible answer to the second has just been published on the website of the United Nations' Sustainable Development Solutions Network (unsdsn.org).

The group proposed ten new Sustainable Development Goals for the next 15 years, each with three draft targets. Some of the language is vague and some of the goals sound like platitudes. But drill down only a little and there are some well-informed suggestions. And some brave ones: a "rapid voluntary reduction of fertility" to address overpopulation, and reform of business accounting and tax regimes among them.

Perhaps most importantly, the draft goals fuse the economic and environmental agendas in a way that the Millennium Development Goals do not. The measure of a national economy, the goals say, is meaningless without a sense of how that bottom line has been achieved, and how it affects people and the planet. The economic currency of gross domestic product, for so long used as a benchmark of a country's performance, could be tweaked to include social indicators and how well a country respects environmental criteria, such as the concept of planetary boundaries that should not be exceeded. The draft report is open for comments until next week. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunq



Driving students into science is a fool's errand

If programmes to bolster STEM education are effective, they distort the labour market; if they aren't, they're a waste of money, argues Colin Macilwain.

The United States spent more than US\$3 billion last year across 209 federal programmes intended to lure young people into careers in science, technology, engineering and mathematics (STEM). The money goes on a plethora of schemes at school, undergraduate and postgraduate levels, all aimed at promoting science and technology, and raising standards of science education.

In a report published on 10 April, Congress's Government Accountability Office (GAO) asked a few pointed questions about why so many potentially overlapping programmes coexist. The same day, the 2014 budget proposal of President Barack Obama's administration suggested consolidating the programmes, but increasing funding.

What no one asked was whether these many activities actually benefit science and engineering, or society as a whole. My answer to both questions is an emphatic 'no'.

Taken individually, of course, these programmes are all very cuddly and wonderful. They are keenly pursued by governments around the world — particularly in countries that fret about their economic competitiveness, such as the United Kingdom and the United States.

But taken together, these schemes — which allocate perhaps \$600 to each child passing through the US education system — constitute bad public policy. Government promotion of science careers ultimately damages science and engineering, by inflating supply and depressing demand for scientists and engineers in the employment market.

Start by asking why no such government-backed programmes exist to pull children into being lawyers or accountants. The obvious answer is that there is no need: young people can see the prospects in these fields for themselves. As a result, places to study these subjects tend to be fiercely competitive. But in many science and engineering disciplines, college places are ten-a-penny after decades of sustained government efforts to render them more attractive.

The dynamic at work here isn't complicated. By cajoling more children to enter science and engineering — as the United Kingdom also does by rigging university-funding rules to provide more support for STEM than other subjects — the state increases STEM student numbers, floods the market with STEM graduates, reduces competition for their services and cuts their wages. And that suits the keenest proponents of STEM education programmes — industrial employers and their legion of lobbyists — absolutely fine.

It's not as if \$3 billion is spent on promoting STEM education each year because US parents

demand it. "I just wish little Mary got the chance to do science at school" is not a phrase, I would submit, that politicians often hear on the doorstep. Nor do universities ask for programmes to encourage more kids to enter undergraduate science.

It is true that some of the larger STEM programmes — notably those at the US National Science Foundation — were backed historically by the Congressional Black Caucus and other organizations in a laudable, but thus far unsuccessful, effort to bring more under-represented minorities into science and engineering.

But the main backing for government intervention in STEM education has come from the business lobby. If I had a dollar for every time I've heard a businessman stand up and bemoan the alleged failure of the education system to produce the science and technology 'skills' that his company requires, I'd be a very rich man.

I have always struggled to recognize the picture these detractors paint. I find most recent science graduates to be positively bursting with both technical knowledge and enthusiasm.

If business people want to harness that enthusiasm, all they have to do is put their hands in their pockets and pay and train newly graduated scientists and engineers properly. It is much easier, of course, for the US National Association of Manufacturers and the British Confederation of British Industry to keep bleating that the state-run school- and university-education systems are 'failing'.

The GAO report on STEM education points out that few of the "complicated patchwork of overlapping programmes" are ever assessed for their effectiveness. Now the Obama administration is proposing, in its 2014 budget, that the existing spread of programmes be consolidated within just three agencies. This proposal sounds eminently reasonable — but is unlikely to happen, given the congressional appropriations process and the sheer impossibility of transferring resources from, say, the National Institutes of Health to the Smithsonian.

Instead of playing political games with the issue, the Obama administration should take a closer look at whether this cluster of activity is worth \$3 billion — especially when essential spending is being cut across the board. The state can't manage or second-guess the labour market, and its efforts to do so are doomed to failure. Government policy should be to tell the education system what most parents tell their own kids: if you love immunology or geophysics, go ahead and do it; if your love is music or investment banking, do that instead. ■

Colin Macilwain writes about science policy from Edinburgh, UK.
e-mail: cfmworldview@gmail.com

**GOVERNMENT
PROMOTION
OF SCIENCE
CAREERS
ULTIMATELY
DAMAGES
SCIENCE AND
ENGINEERING.**

NATURE.COM
Discuss this article
online at:
go.nature.com/bnawjp

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

NEUROSCIENCE

Cell transplants stem seizures

Stem-cell therapy can reduce seizures in epileptic mice.

Some forms of epilepsy are thought to be caused by dysfunctional cells in the hippocampus region of the brain. The affected cells, called inhibitory interneurons, help to regulate neural circuits. Robert Hunt, Scott Baraban and their colleagues at the University of California, San Francisco, harvested progenitors of inhibitory neurons from embryonic mice and injected them into the hippocampi of epileptic adult mice.

In the brain, the cells matured into functioning inhibitory interneurons. Mice that received cell transplants had many fewer seizures than untreated mice. The transplants also reversed some behavioural problems that epileptic mice often develop, such as hyperactivity and poor spatial learning.

Nature Neurosci. <http://dx.doi.org/10.1038/nn.3392> (2013)

ADDICTION RESEARCH

Drug vaccine blocks rat relapse

A vaccine stops heroin-addicted rats from seeking the drug.

Earlier vaccines against heroin have faltered because the antibodies that they induce are easily overwhelmed by increased doses of the drug. Joel Schlosburg and his colleagues at the Scripps Research Institute in La Jolla, California, have created a three-in-one vaccine: the antibodies it triggers target not only heroin, but also two of the drug's breakdown products. Active molecules are bound up by the antibodies

before they can slip from the bloodstream into the brain. Vaccinated rats given heroin did not exhibit the compulsive drug-seeking behaviours seen in unvaccinated rats.

Proc. Natl Acad. Sci. USA <http://dx.doi.org/10.1073/pnas.1219159110> (2013)

CLIMATE SCIENCE

More cyclones for Hawaiian Islands

Climate change could double the number of tropical cyclones that storm into the Hawaiian Islands.

Hiroiyuki Murakami, then at the Meteorological Research

Institute in Tsukuba, Japan, and his colleagues simulated tropical-cyclone patterns using several versions of a weather model, as well as predictions of sea-surface warming. The authors used the simulations, which they verified against modern observations, to predict that the frequency of cyclones reaching the Hawaiian Islands during 2075–99 will be about double that in 1979–2003.

The change is due mainly to the way in which cyclones that arise off the west coast of Mexico are expected to move across the Pacific Ocean as the climate warms. Although fewer storms will form, more

will take a path towards the islands and have enough power to reach them.

Nature Clim. Change <http://dx.doi.org/10.1038/nclimate1890> (2013)

MICROBIOLOGY

Fast life in the urinary tract

Bacteria responsible for recurrent urinary-tract infections migrate from the gut.

These infections, which affect up to one-quarter of women, are caused mostly by strains of the bacterium *Escherichia coli* that are thought to give up their



BLACKSMITH INST.

EPIDEMIOLOGY

Unappreciated toll of toxic sites

Toxic-waste sites pose as big a health threat as malaria in some developing nations.

Kevin Chatham-Stephens at the Icahn School of Medicine at Mount Sinai in New York and his colleagues produced a systematic assessment of the disease burden of toxic-waste sites in India (pictured), Indonesia and the Philippines.

The team used site visits, sampling, interviews and a global database of dumps to estimate that, in 2010, 8.6 million people were at risk

of exposure to industrial pollutants — mainly lead and hexavalent chromium — at 373 sites in these nations. The authors calculated that the toxic exposures from just the surveyed sites resulted in the loss of more than 800,000 years of healthy life owing to ill-health and death in these countries, a toll about half that of air pollution and slightly more than that of malaria.

Environ. Health Perspect. <http://dx.doi.org/10.1289/ehp.1206127> (2013)

adaptations to life in the gut for ones that allow them to thrive in the urethra and bladder. A team led by Jeffrey Gordon at Washington University in St. Louis, Missouri, compared the genomes of dozens of *E. coli* strains from the faeces and urine of four women with recurrent infections.

In two of the women, new but matching *E. coli* strains emerged in both the gut and the urinary tract. Follow-up experiments in mice showed that the same strain could survive at both sites without the expected loss of fitness. Strategies to fight recurrent infections might need to be reconsidered, the authors say. *Sci. Transl. Med.* 5, 184ra60 (2013)

ANIMAL BEHAVIOUR

Tree-loving lemur digs hole in winter

Lemurs normally live in trees, but researchers have discovered that at least two species hibernate underground.

Marina Blanco at the Duke Lemur Center in Durham, North Carolina, and her colleagues unearthed two species of eastern dwarf lemur (*Cheirogaleus sibreei* and *C. crossleyi*, pictured) in Tsinjoarivo, a high-altitude rainforest in central-eastern Madagascar.

By tagging 12 animals with radiotransmitters, the team found that the lemurs dig underground burrows in which they hibernate for several months between April and September. By contrast, the fat-tailed dwarf lemur (*C. medius*),

the only other primate species known to hibernate annually, sleeps in holes in tree trunks. Being buried beneath an insulating layer of leaves and roots might help the eastern dwarf lemurs to maintain stable body temperatures amid large daily swings in air temperature, which often dip to 5 °C in the cold season.

Sci. Rep. 3, 1768 (2013)

PALAEOLOGY

Flow sorting for fossil pollen

Unconventional use of a sorting technology could help palaeontologists to date lake sediments. This could enable better reconstruction of past environments and climates.

Pollen grains, which can be used in carbon dating, are often the only organic matter found in abundance in lake sediments. However, isolating the grains is difficult, often making their use in carbon-dating impractical. Richard Jones at the University of Exeter, UK, and his colleagues adapted a biomedical technique known as flow cytometry, which sorts labelled cells on the basis of how they fluoresce and scatter light. Pollen grains are naturally fluorescent, so could be separated from contaminants by size and shape.

In just four hours, the technique collected 2.75 million fossil grains — an amount large enough to date a sample.

J. Quart. Sci. 28, 229–236 (2013)

CELL BIOLOGY

Sperm control DNA breaks

When dividing to produce mammalian sex cells, cells coordinate how DNA breaks occur.

The cell-division process that forms these sex cells, which contain one set of chromosomes instead of the normal two, involves the swapping of bits of genetic material between matching

COMMUNITY CHOICE

The most viewed papers in science

MATERIALS SCIENCE

Holes help supercapacitor

HIGHLY READ
on www.wiley.com in March

A graphene-based material can store energy at seven times the density of commercial carbon-based products.

Supercapacitors — which could be used to power electric vehicles — recharge and release energy faster than batteries, but their energy density is lower. Theoretically, graphene — an unusual material made of atomically thick layers of carbon — is a good candidate for a supercapacitor, but its carbon layers tend to stack together, degrading its performance. Pei Kang Shen and his colleagues at Sun Yat-sen University in Guangzhou, China, created a porous graphene-based structure using an acrylic resin as a template. The researchers rinsed the resin with a nickel acetate solution followed by acid to produce a network of graphene layers with interconnected pores. When the material was filled with an electrolyte solution, it showed stable performance and had an energy density comparable to that of a lead-acid battery.

Adv. Mater. 25, 2474–2480 (2013)

pairs of chromosomes. This process can involve hundreds of breaks in the double-stranded DNA. A team led by Maria Jasin and Scott Keeney at the Memorial Sloan-Kettering Cancer Center in New York found that in male mice engineered to have reduced double-stranded breaks, chromosome pairs failed to exchange DNA properly. The resulting ‘chromosome tangles’ continued to accumulate breaks as cell division progressed, suggesting that a feedback mechanism stalls breaks as chromosomes successfully intermingle. The cells of the engineered mice eventually halted division, and the mice were left sterile.

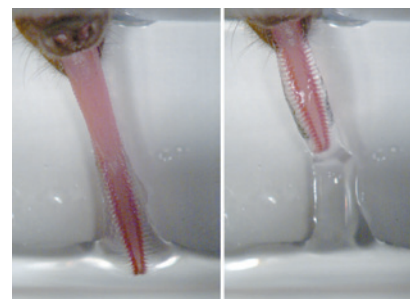
Genes Dev. 27, 873–886 (2013)

ZOOLOGY

Tongue spikes snare nectar

Hovering bats use barbed tongues to snare nectar from flowers.

Cally Harper and her colleagues at Brown University in Providence, Rhode Island, used a high-speed video camera and post-mortem



analyses to understand the mechanics of the eponymous organ of the Pallas's long-tongued bat (*Glossophaga soricina*). The tongue (pictured) can extend up to 4 centimetres, about twice the length of the bat's head. The researchers found that as bats thrust out their tongues, muscles force blood into rows of tiny bristles that usually lie flat along the tip of the tongue. The bristles become erect in 0.04 seconds, trapping nectar between them as the bat retracts its tongue.

Proc. Natl Acad. Sci. USA
<http://dx.doi.org/10.1073/pnas.1222726110> (2013)

NATURE.COM

For the latest research published by Nature visit:

www.nature.com/latestresearch

CALLY HARPER

GEOFF TRINDER/ARDEA.COM



SEVEN DAYS

The news in brief

EVENTS

Polio in Somalia

Somalia has recorded its first case of wild poliovirus since March 2007, the World Health Organization said on 11 May. A 32-month-old girl living near Mogadishu has been paralysed by the virus. There have been no polio immunizations in some parts of the country since 2009. As *Nature* went to press, genetic testing was under way to determine the virus's origin. The only other countries that have recorded wild poliovirus cases in 2013 are Pakistan, Afghanistan and Nigeria, the three countries that have never managed to interrupt its transmission. See go.nature.com/x43baq for more.

POLICY

Online data tool

US President Barack Obama on 9 May unveiled a public tool for online searches and analysis of information from more than 9,000 government data sets on health, climate science and the environment. Such data could help scientists to link extreme weather events with common illnesses or infectious diseases, for example. The project will be coordinated across six government agencies, including the Centers for Disease Control and Prevention and the US Geological Survey. Its launch accompanied the announcement of a policy that would make all government-generated data more accessible by computer.

Radiation warnings

Lawmakers in San Francisco, California, agreed on 7 May to strike down an ordinance that required retailers to warn consumers about allegedly dangerous radiation from mobile

phones. Implementation of the 2011 city law — the first of its kind in the United States — had been blocked by legal challenges from a communications-industry group. City leaders voted to settle with the group by revoking the law. In exchange, the industry association agreed to waive its claim to roughly US\$500,000 in legal fees that it might have collected in a court victory.

Research revamp

Canada is changing the focus of its National Research Council (NRC) from broad-ranging, independent science to commercial research directed by private industry. The overhaul of the 97-year-old government agency, revealed on 7 May, is aimed

at spurring technology innovation and economic growth. The NRC institutes — budgeted at Can\$900 million (US\$889 million) in 2013 — will be reorganized into 12 business units focused on 5 areas of research: security, manufacturing, infrastructure, health costs, and natural resources and the environment.

RESEARCH

Neuroscience goal

Researchers met for the first time to begin planning the US National Science Foundation's part in the Brain Research through Advancing Innovative Neurotechnologies initiative — an ambitious bid to understand the brain's signals by simultaneously recording from unprecedented numbers

of neurons. More than 150 scientists from such fields as neuroscience, engineering and computer science attended the conference in Arlington, Virginia, which ended on 7 May. Also present were scientists representing the US National Institutes of Health — one of the initiative's two other government-agency partners. See go.nature.com/xgsrwa for more.

FACILITIES

Campus upset

University College London (UCL) announced on 7 May that it has abandoned plans to build a £1-billion (US\$1.54-billion) campus on a site in east London that is currently a residential estate. The university had been in



LAURA LEON/POLARIS/EYEVINE

Endangered ecosystems get listed

The International Union for Conservation of Nature (IUCN) on 8 May updated the criteria for its Red List of Ecosystems (D. A. Keith *et al.* *PLoS ONE* 8, e62111; 2013). Based on its successful Red List of Threatened Species, which assesses risk of extinction, the new list ranks ecosystems using factors such as rate of shrinkage, disruption to wildlife and risk

of ecosystem collapse. The paper highlights the Aral Sea in central Asia as an example of a collapsed ecosystem. River-diversion projects have caused the body of water, once the fourth-largest lake in the world, to lose about 90% of its volume in the past 50 years, stranding fishing boats (pictured) and depleting native wildlife.

SCOTT J. FERRELL/CONGRESSIONAL QUARTERLY/GETTY IMAGES

talks with the London Borough of Newham to develop teaching and research facilities on the estate, some 9 kilometres from UCL's base in central London. UCL says that it could not reach agreement with local authorities over the redevelopment, but insists that it is committed to establishing quarters in the area.

FUNDING

NIH budget blues

The US National Institutes of Health (NIH) announced on 8 May a final 2013 budget of US\$29.15 billion, down from \$30.7 billion in 2012. The revised number reflects the effect of a government spending cut called the sequester, which was implemented in March. As a result, the agency said that it will fund 34,902 research grants — 3.7% fewer than in 2012. Grant recipients awaiting instalments on grants won in previous years will see these funded “at levels below those indicated on the Notice of Award”, the agency said.

Smith under fire

US congressman Lamar Smith (Republican, Texas) is facing mounting opposition to draft legislation that would alter the way research is funded by the National Science Foundation (NSF). Under the proposal drafted by Smith (**pictured**),



who leads the House science committee, all NSF-funded research would be required to expressly benefit national interests and to undergo additional certification after peer review. In two letters sent to Smith on 8 May, 24 scientific leaders, including former NSF directors and assistant directors, and former chairmen of the National Science Board, requested that Smith retract the legislation, saying that it would hinder scientific progress.

PEOPLE

Stalled nominee

Eight Republican senators boycotted a 9 May vote on the nomination of Gina McCarthy as administrator of the US Environmental Protection Agency (EPA), effectively blocking the process. McCarthy would oversee US efforts to implement regulations targeting global warming

and other environmental issues. The senators, from the Committee on Environment and Public Works, suggest that they will vote once the EPA has responded to their questions over the proposed regulations. Senate Democrats have rescheduled the vote for 16 May.

Hawking row

Physicist Stephen Hawking of the University of Cambridge, UK, has pulled out of a high-profile conference in Israel. Hawking said his decision, which became public last week, was made to “respect the boycott” of Israel over its actions in the occupied Palestinian territories. His move has rekindled fierce debate among proponents and opponents of the boycott. Hawking's withdrawal has drawn fire from both the conference organizers and the Israel Academy of Science and Humanities in Jerusalem (see page 299 for more).

BUSINESS

Shale-gas exit

Two energy firms — Talisman Energy, headquartered in Calgary, Canada, and Marathon Oil, based in Houston, Texas — decided last week to cease shale-gas exploration activities in Poland. The pull-out raises further questions about Poland's shale-gas production,

COMING UP

18–21 MAY

The American Society for Microbiology holds its annual general meeting in Denver, Colorado, with scientific discussions including threats from the avian influenza viruses H7N9 and H5N1.

gm.asm.org

21–23 MAY

The Pasteur Institute in Paris hosts an international symposium on HIV research, marking 30 years since the first reports of a retrovirus associated with AIDS.

www.30yearshiv.org

which is yet to begin. In a 2011 review, the US Energy Information Administration estimated that Poland had 5.3 trillion cubic metres of shale gas, but a 2012 study by the Polish Geological Institute revised this figure to less than 800 billion cubic metres.

Transgene patents

The US Supreme Court ruled on 13 May that a farmer had violated intellectual-property laws by planting genetically modified soya beans without buying the goods from the crop's patent-holder, agricultural technology giant Monsanto, based in St Louis, Missouri. Indiana farmer Vernon Bowman argued that Monsanto's patents did not apply to seeds he purchased from a grain elevator (storage tower) that contained a mixture of surplus crops, including Monsanto's herbicide-resistant soya beans. The court disagreed, saying that US patent law “provides no haven” for propagating crops from such seeds. See go.nature.com/uil764 for more.

► **NATURE.COM**

For daily news updates see:
www.nature.com/news

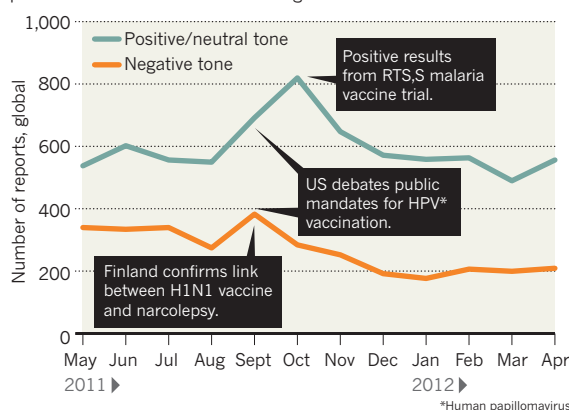
SOURCE: HEIDI LARSON/LANCET INFECT. DIS.

TREND WATCH

An online media surveillance tool can track concerns and rumours about vaccines as they occur in real time. The system, described in a 13 May paper (H. J. Larson *et al.* *Lancet Infect. Dis.* <http://doi.org/mhm>; 2013), picks up positive and negative sentiments, both worldwide (see chart) and for particular events in specific countries. Heidi Larson of the London School of Hygiene & Tropical Medicine hopes that real-time monitoring will help health officials to know where to focus if trust in vaccines falls.

VACCINES GET A VOTE OF CONFIDENCE

Negative reports about vaccines were outweighed about 2:1 by positive or neutral sentiments during 2011 and 2012.



NEWS IN FOCUS

ASTRONOMY All eyes on the Milky Way's turbulent centre **p.296**

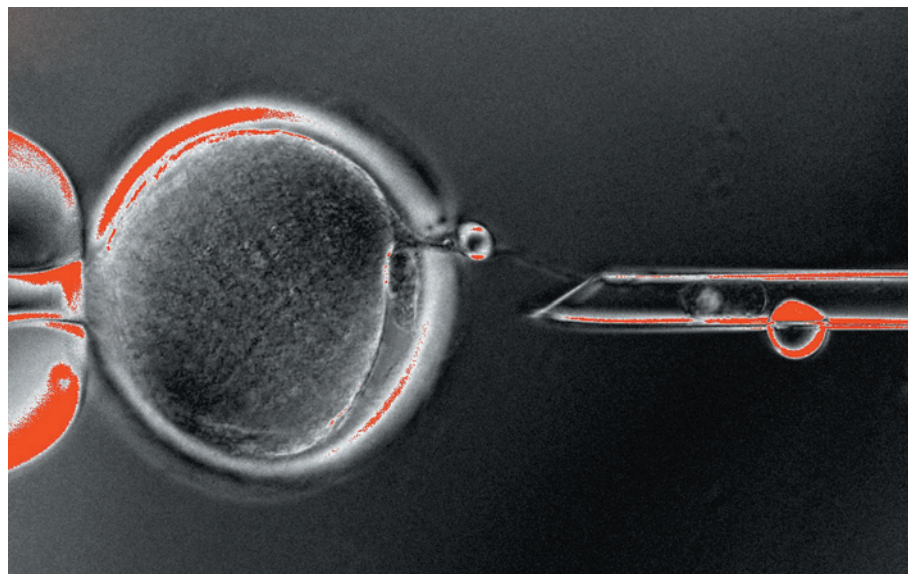
GENOMICS Chinese project probes the genes of genius **p.297**

POLITICS Stephen Hawking provokes war of words over Israel boycott **p.299**

CONSERVATION A rain of death for invasive Galapagos rats **p.306**



OHSU PHOTOS



almost any cell type in the human body.

Many scientists have tried to create human SCNT cell lines; none had succeeded until now. Most infamously, Woo Suk Hwang of Seoul National University in South Korea used hundreds of human eggs to report two successes, in 2004 and 2005. Both turned out to be fabricated. Other researchers made some headway. Mitalipov created SCNT lines in monkeys² in 2007. And Dieter Egli, a regenerative medicine specialist at the New York Stem Cell Foundation, successfully produced human SCNT lines³, but only when the egg's nucleus was left in the cell. As a result, the cells had abnormal numbers of chromosomes, limiting their use.

MONKEYING AROUND

Mitalipov and his group began work on their new study last September, using eggs from young donors recruited through a university advertising campaign. In December, after some false starts, cells from four cloned embryos that Mitalipov had engineered began to grow. "It looks like colonies, it looks like colonies," he kept thinking. Masahito Tachibana, a fertility specialist from Sendai, Japan, who is finishing a 5-year stint in Mitalipov's laboratory, nervously sectioned the 1-millimetre-wide clumps of cells and transferred them to new culture plates, where they continued to grow — evidence of success. Mitalipov cancelled his holiday plans. "I was happy to spend Christmas culturing cells," he says. "My family understood."

The success came through minor technical tweaks. The researchers used inactivated Sendai virus (known to induce fusion of cells) to unite the egg and body cells, and an electric jolt to activate embryo development. When their first attempts produced six blastocysts but no stable cell lines, they added caffeine, which protects the egg from premature activation.

None of these techniques is new, but the researchers tested them in various combinations in more than 1,000 monkey eggs before moving on to human cells. "They made the right improvements to the protocol," says Egli. "It's big news. It's convincing. I believe it."

The experiments took only a few months, Mitalipov says. "People say, you did it in monkeys in 2007. Why did it take six years in humans?" Most of the time, he says, was spent navigating US regulations on embryo research.

The researchers carried out a battery of tests to prove that their SCNT cells could form various cell types, including heart cells that

Seeing double: human embryonic stem cells have finally been made using cloning techniques.

REGENERATIVE MEDICINE

Human stem cells created by cloning

Breakthrough sets up showdown with induced adult lines.

BY DAVID CYRANOSKI

It was hailed some 15 years ago as the great hope for a biomedical revolution: the use of cloning techniques to create perfectly matched tissues that would someday cure ailments ranging from diabetes to Parkinson's disease. Since then, the approach has been enveloped in ethical debate, tainted by fraud and, in recent years, overshadowed by a competing technology. Most groups gave up long ago on the finicky core method — production of patient-specific embryonic stem cells (ESCs) from cloning. A quieter debate followed: do we still need 'therapeutic' cloning?

A paper published this week¹ by Shoukhrat Mitalipov, a reproductive biology specialist at the Oregon Health and Science University in Beaverton, and his colleagues is sure to

rekindle that debate. Mitalipov and his team have finally created patient-specific ESCs through cloning, and they are keen to prove that the technology is worth pursuing.

Therapeutic cloning, or somatic-cell nuclear transfer (SCNT), begins with the same process used to create Dolly, the famous cloned sheep, in 1996. A donor cell from a body tissue such as skin is fused with an unfertilized egg from which the nucleus has been removed. The egg 'reprograms' the DNA in the donor cell to an embryonic state and divides until it has reached the early, blastocyst stage. The cells are then harvested and cultured to create a stable cell line that is genetically matched to the donor and that can become

NATURE.COM

For more on induced pluripotent stem cells, see: go.nature.com/otnzri

► are able to contract spontaneously.

Their first cell lines were created using fetal skin cells; others were derived using donor cells from an 8-month-old patient with a rare metabolic disorder called Leigh syndrome, to prove that ESCs could be made from more mature donor cells. The technique does not require prohibitive numbers of eggs: it took 15 from one donor to produce one cell line and 5 from a different donor to make another. “The efficiency was the most impressive thing,” says George Daley, a stem-cell expert at Children’s Hospital Boston in Massachusetts.

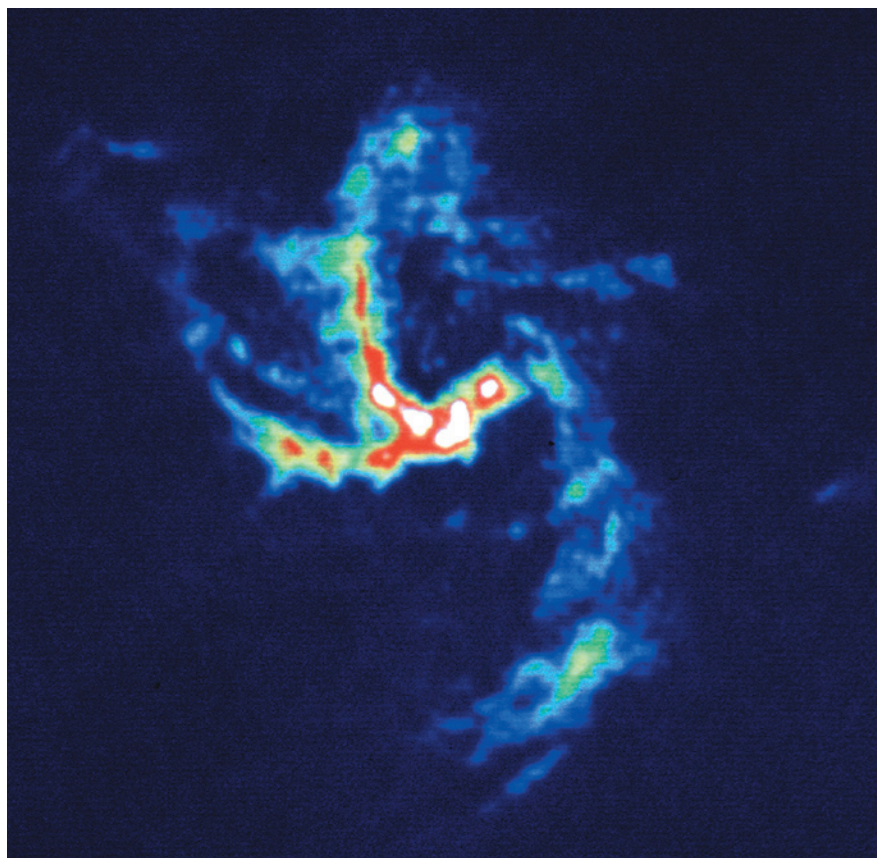
Such improvements might be necessary to convince people that SCNT research is still worthwhile. Egg donors for the experiment received US\$3,000–7,000 in compensation. This is expensive and, according to some bioethicists, risks creating an organ trade that preys on the poor. Because the technique requires the destruction of embryos, funds from the US National Institutes of Health (NIH) cannot be used to make or study SCNT-derived cell lines, hampering further clinical research. (Mitalipov maintains a separate laboratory for NIH-funded research.)

Public fears that the technology might be used to create human clones are another sticking point. The research might spark “cloning hysteria” that opponents of stem-cell research could capitalize on, says Bernard Siegel, executive director of the Genetics Policy Institute in Palm Beach, Florida. But Mitalipov has tried without success for more than a decade to produce a monkey by cloning. Tachibana says that an upcoming publication will explain why reproductive cloning of humans is not possible using their SCNT technique.

Still, Daley and most other stem-cell researchers have shifted to another method for creating genetically matched, patient-specific cell lines: reprogramming adult cells to an embryonic state to produce induced pluripotent stem (iPS) cells. First reported in 2006, the technique does not involve eggs, cloning or destruction of embryos⁴. “Honestly, the most surprising thing [about this paper] is that somebody is still doing human [SCNT] in the era of iPS cells,” says Miodrag Stojkovic, who studies iPS cells for regenerative medicine and runs a fertility clinic in Leskovac, Serbia.

But Stojkovic, like others, awaits the results of head-to-head comparisons between iPS and SCNT cells. Some research has shown that iPS cells are not completely reprogrammed and that stem cells derived from SCNT are more like embryonic stem cells derived from *in vitro* fertilization. Mitalipov and Tachibana are now conducting a study to compare iPS cells and SCNT cells derived from the same donor cell. “These results,” says Daley, “will be fascinating.” ■

1. Tachibana, M. *et al.* *Cell* <http://dx.doi.org/10.1016/j.cell.2013.05.006> (2013).
2. Byrne, J. A. *et al.* *Nature* **450**, 497–502 (2007).
3. Noggle, S. *et al.* *Nature* **478**, 70–75 (2011).
4. Takahashi, K. & Yamanaka, S. *Cell* **126**, 663–676 (2006).



Gas clouds swirl around the black hole at the Milky Way’s centre (seen here in radio emission).

ASTRONOMY

Magnetar found at giant black hole

Magnetized neutron star could test Einstein’s theory.

BY EUGENIE SAMUEL REICH

Dale Frail couldn’t resist the prospect of watching a black hole swallow its prey. Frail, who is in charge of the Very Large Array (VLA) of radio telescopes near Socorro in New Mexico, had seen a report last month about a long-lived X-ray flare emanating from the centre of the Milky Way, home to a supermassive black hole called Sagittarius A* (Sgr A*). Astronomers were speculating that the flare might be a sign that a gas cloud they had been tracking had begun its death spiral into the black hole.

Frail was sceptical. The cloud’s death was not expected until between September this year and March 2014. But Frail did not want to risk missing the action. Within hours of seeing the report, he had trained the VLA’s radio dishes on the scene, only to find

nothing remarkable. Frail was puzzled. If the flare wasn’t the arrival of the gas cloud, what was it?

An answer soon came from other telescopes watching the drama at the centre of the Galaxy: the flare was coming from a magnetar, a highly magnetized kind of pulsar, or rotating neutron star. Its position near Sgr A* makes it a precious find. The magnetar’s regular radio pulses could be used to measure the warping of space-time near the monster black hole and to test predictions of Einstein’s general theory of relativity.

“There’s huge interest in finding pulsars around supermassive black holes, and this is the first example,” says Geoffrey Bower, a radioastronomer at the University of

➔ **NATURE.COM**

For more on the Milky Way’s black hole, see:

go.nature.com/apzhzf

California, Berkeley, who carried out the recent VLA observations. "There's a lot we can learn from this."

The magnetar's accidental discovery is a by-product of astronomers' excitement about the arrival of the gas cloud, dubbed G2. The cloud, which is about three times the mass of Earth, was first spotted near Sgr A* in 2012 (and was later found in 2002 data). Its arrival would deliver insight into how objects accrete into the swirling disk of material around a black hole, as well as offering the first chance for astronomers to measure the time that it takes for objects to be captured and swallowed up.

Every flicker of emissions from Sgr A* sparks a flurry of speculation, intensifying the usual cycle of observation and coordinated follow-up that characterizes high-energy astronomy. Many telescope directors are scheduling additional monitoring of the Galactic Centre. The VLA, for example, is already scanning radio frequencies around Sgr A* every two months, and will do so every month once G2 arrives.

"I don't think there was ever such a large camp of telescopes looking at the Galactic Centre," says Stefan Gillessen, an astronomer at the Max Planck Institute for Extraterrestrial Physics in Garching, Germany, who last year reported that G2 was on course for Sgr A* (S. Gillessen *et al. Nature* **481**, 51–54; 2012).

The observing frenzy is likely to jeopardize regularly scheduled observations, and some astronomers worry that the pay-off may be disappointing. Andrea Ghez, an astronomer

studying G2 at the University of California, Los Angeles, says that her infrared observations at the Keck telescopes on Mauna Kea in Hawaii suggest that G2 may not be a gas cloud, but rather a star surrounded by gas. If she is right, Sgr A* may swallow some of the gas, but the star itself would have enough momentum to escape the black hole's grasp. The result would be a Galactic fizzle rather than fireworks. "I'm just worried that this is overblown," she says.

Yet the focus on the Galactic Centre has paid off unexpectedly. The reported X-ray flare prompted Fiona Harrison, an astronomer who leads NASA's Nuclear Spectroscopic Telescope Array (NuSTAR) satellite, to train the X-ray space telescope on Sgr A*. Like Frail, Harrison doubted that the flare had anything to do with G2, but she thought that NuSTAR's ability to time events precisely might clarify the situation. Sure enough, on 26 April, NuSTAR detected an X-ray signal blinking every 3.76 seconds — the signal of a pulsar spinning at that rate. Then, on 29 April, the Chandra X-ray Observatory localized the magnetar to at least 0.12 parsecs (0.38 light years) from the black hole — close, but much farther away than G2.

On 4 May, observations by NuSTAR and Swift, the NASA space telescope that spotted the original flare, identified the object conclusively as a magnetar by showing a gradual slowing in its spin rate. This slowing is consistent with the presence of the high magnetic field that enables the star to radiate energy faster than a normal pulsar.

Fourteen magnetars, including this one, have been found in the general region of the Galactic Centre. This high number supports the idea that magnetars tend to form from the death throes of the bright, heavy stars that are common there, says Chryssa Kouveliotou, a magnetar expert at NASA's Marshall Space Flight Center in Huntsville, Alabama.

The finding also hands astronomers a tool for studying conditions near supermassive black holes. According to Einstein's general theory of relativity, clocks in high gravitational fields run slow, as a result of the warping of space-time. So if the mag-

netar is following an elliptical orbit around the black hole, its clock-like spin rate should speed up and slow down as its distance from the black hole varies — an effect that, with luck, could be disentangled from the gradual slow-down caused by the magnetar's magnetic field.

Frail suggests that the magnetar's discovery shows that all the excitement over G2 is justified. Astronomers watching G2 say that the cloud itself could deliver plenty more excitement when it finally reaches the Galactic Centre. They are undaunted by the fact that one of the first apparent signs of its arrival turned out to be something completely different. "I wish all of our failed experiments were that good," says Frail. ■

"I don't think there was ever such a large camp of telescopes looking at the Galactic Centre."

GENOMICS

Chinese project probes the genetics of genius

Bid to unravel the secrets of brainpower faces scepticism.

BY ED YONG

The US adolescents who signed up for the Study of Mathematically Precocious Youth (SMPY) in the 1970s were the smartest of the smart, with mathematical and verbal-reasoning skills within the top 1% of the population. Now, researchers at BGI (formerly the Beijing Genomics Institute) in Shenzhen, China, the largest gene-sequencing facility in the world, are searching for the quirks of DNA that may contribute to such gifts. Plunging into an area that is littered with failures and riven with controversy, the researchers are scouring the genomes of 1,600 of these high-fliers in an ambitious project to find the first common

genetic variants associated with human intelligence.

The project, which was launched in August 2012 and is slated to begin data analysis in the next few months, has spawned wild accusations of eugenics plots, as well as more measured objections by social scientists who view such research as a distraction from pressing societal issues. Some geneticists, however, take issue with the study for a different reason. They say that it is highly unlikely to find anything of interest — because the sample size is too small and intelligence is too complex.

Earlier large studies with the same goal have failed. But scientists from BGI's Cognitive Genomics group hope that their super-smart

sample will give them an edge, because it should be enriched with bits of DNA that confer effects on intelligence. "An exceptional person gets you an order of magnitude more statistical power than if you took random people from the population — I'd say we have a fighting chance," says Stephen Hsu, a theoretical physicist from Michigan State University in East Lansing, who acts as a scientific adviser to BGI and is one of the project's leaders.

"If they think they're likely to get much useful data out of this study, they're almost certainly wrong," says Daniel MacArthur, a geneticist at Massachusetts General Hospital in Boston. He is not against intelligence ►



Candidates for the Study of Mathematically Precocious Youth in 1983. Study data are in use at China's BGI.

► studies in principle, despite the visceral reactions they provoke in some people. “Studying intelligence is useful for understanding cognitive function, or diseases” that affect it, he says. But he questions whether the study will work.

Intelligence has a substantial but mysterious genetic component¹. Studies in twins indicate that genetic factors should explain significantly more than half of the variation in adult general intelligence — the abstract quantity measured in IQ tests. This, in turn, correlates well with attributes such as academic achievement and income. Although geneticists have identified hundreds of genes in which a single mutation can lead to developmental difficulties, the common genetic variants that lie behind normal diversity in intelligence remain elusive. It is widely assumed that, as is the case for traits such as height, there are thousands of such variants, the small effects of which combine to influence mental abilities.

So far, the quest to identify intelligence-linked variants has thrown up only null results and false positives², even after large, well-designed studies. In 2010, a team led by Robert Plomin, a behavioural geneticist at King's College London, failed to find a single intelligence-associated variant, even after

examining more than 350,000 variations in single DNA letters, or SNPs, across the genomes of 7,900 children³.

After this, Plomin switched his strategy to focus on only the brightest minds. He collected DNA samples from 2,000 of the SMPY's recruits, whose average IQ is above 150 — surpassing the average of Nobel laureates and putting them three standard deviations above the general population's mean score of 100. “In the earlier study, I bet we didn't have more than two or three people with an IQ that high,” says Plomin, who has been studying the heritability of intelligence since the 1970s.

At the same time, BGI had hit on a similar strategy and was trying to recruit the brightest teenagers for a new study. Although Hsu is not a geneticist himself, he has a long-standing interest in intelligence. He suggested that BGI should recruit gifted individuals from science and mathematics olympiads — international competitions for school pupils — but when that approach failed, Hsu set up a website to attract volunteers of ‘high cognitive ability’.

“If they think they're likely to get much useful data out of this study, they're almost certainly wrong.”

Then he heard about Plomin's sample. The two struck up a partnership: Plomin supplied DNA samples from 1,600 SMPY recruits, and Hsu added samples from more than 500 people recruited — albeit less selectively — through his website, which asked potential donors to submit evidence of their brain power in the form of impressive olympiad results, high scores on standardized tests or diplomas from leading universities. “We're at the bleeding edge,” says Hsu. “No one has a sample like this.”

BGI is sequencing the genomes at low coverage — scanning each DNA letter an average of only four or five times — and comparing them with what Hsu says will be in excess of 4,000 controls from the UK10K project, which aims to sequence the genomes of 10,000 UK residents. Because these people were not selected for intelligence, they should display a range of IQs. The researchers at BGI will begin by using commercially available tools to look at common SNPs before moving on to rarer ones, and to variations in the number of copies of different DNA segments.

Plomin is optimistic about the study's chances: “I'm not expecting to find a lot but I'd be happy with one solid finding,” he says. But MacArthur notes that searches for genes connected with schizophrenia have had similar sample sizes and compared people at the far end of the risk spectrum (extremely likely to get the disease) with those in the middle (of average risk). Yet the efforts came up empty-handed. If the genetics of intelligence are similar to those of schizophrenia or height, MacArthur says, the team needs at least 10,000 cases and 10,000 controls.

Sample size could be an even greater handicap if it turns out that intelligence is not actually the result of multiple variants, each with a small effect. Plomin likens a super-smart brain to a finely tuned race car: “You need many positive things going for you and you can't afford many negatives,” he says.

But this may not be true, says Kevin Mitchell, a neurogeneticist from Trinity College Dublin. “Intellectual disability can be caused by a mutation in any one of a large number of genes — we don't know if that is true for exceptional genius,” he says. If rare mutations produce extremes in intelligence just as they do extremes in height, the BGI study will be unlikely to find them.



**MORE
ONLINE**

TOP STORY



Blood hormone relieves age-related heart stiffening and thickening in mice go.nature.com/eosuug

MORE NEWS

- Crowdsourcing project set to locate world's power plants go.nature.com/eoasws
- Official data contradict Spanish government's claims on science funding go.nature.com/3cl6k6
- Common origin for Earth and Moon water go.nature.com/mi2urw

VIDEO OF THE WEEK



Artificial volcanoes offer insight into explosions go.nature.com/yvcen9

Both Plomin and Hsu are passionate enough to take a shot, although their goals differ. Hsu is focused on the genetic basis of extreme intelligence. "My primary interest is why Einstein or Hawking is different from a normal person," he says. Plomin is sequencing high-performers as a way of homing in on genes that affect intelligence in the broader population. If enough of these are discovered, he thinks that it may be possible to predict someone's intelligence from an early age, and to offer help to children who are at risk of learning disabilities.

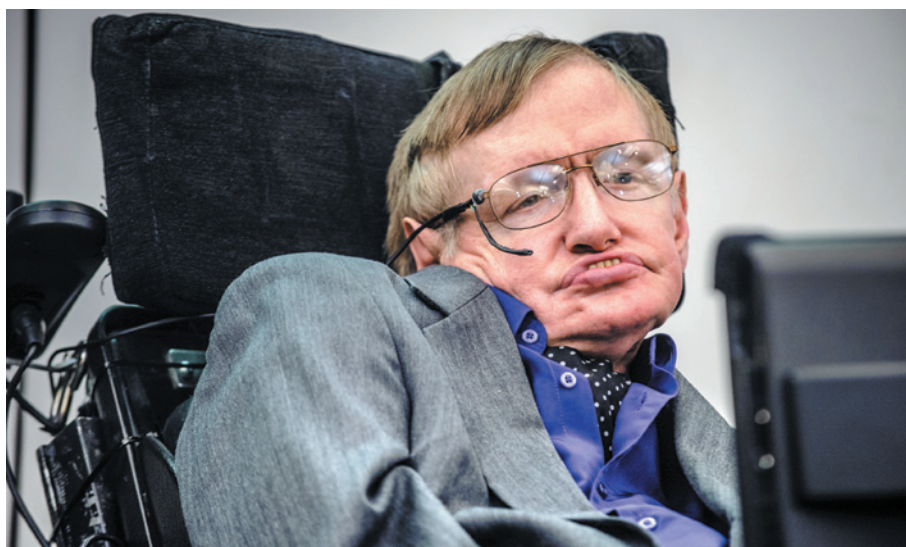
Publicity around the project has spawned some extreme reactions. An article published in March entitled 'China is Engineering Genius Babies' in the US arts and culture magazine *VICE* branded the study "a state-endorsed genetic-engineering project" that will allow parents to predict the IQs of embryos and selectively breed ever-smarter children. ("That's nuts," says Hsu.) "Intelligence does push a lot of buttons. It's like waving a red flag to a bull," says Plomin. He argues that there is nothing wrong with using genetic information as the basis of educational interventions. "I'm interested in predicting learning problems early rather than waiting until kids get to school and then fail," he says.

But Paul Martin, a sociologist at the University of Sheffield, UK, is surprised that geneticists are still pursuing this line of research. "I think most people would say that's the wrong paradigm, when most educational research suggests that social factors are incredibly important," he says. "Strategically, this seems like something of a throwback."

More controversy is likely to flare when or if an intelligence factor emerges from the project. BGI is halfway through its sequencing, but it is unclear when findings will be released. "You might not see anything for a year," says Hsu; it depends on how long the analysis takes. In the meantime, as the cost of sequencing continues to fall, larger sample sizes may soon become available for studies of this kind. "In 2009, we thought it'd be impossible to have a schizophrenia study with 20,000 cases," says Peter Visscher, a geneticist from the University of Queensland in Brisbane, Australia. "Now, the international community is up to 40,000."

Still, he adds that the idea of predicting intelligence from a DNA sample is fanciful. "Even for human height, where you have samples of hundreds of thousands, the prediction you'd get for a newborn person isn't very accurate," he says. "That will be true for IQ for a long time to come." ■

1. Deary, I. J., Johnson, W. & Houlihan, L. M. *Hum. Genet.* **126**, 215–232 (2009).
2. Chabris, C. F. *et al. Psychol. Sci.* **23**, 1314–1323 (2012).
3. Davis, O. S. *et al. Behav. Genet.* **40**, 759–767 (2010).



PETER VAN DEN BERG/PHOTOSHOT

Stephen Hawking joins a list of scientists who have boycotted Israeli institutions and events.

ACADEMIC BOYCOTTS

Hawking decision fuels Israel debate

Physicist's stance a 'turning point', claim supporters.

BY DANIEL CRESSEY

A decision by one of the world's most famous physicists to withdraw from a conference due to be held in Israel next month has rekindled fierce debate in the United Kingdom over academic boycotts that are intended to protest against Israel's actions in the occupied Palestinian territories.

The decision, by Stephen Hawking of the University of Cambridge, UK, comes hot on the heels of an official ruling against a UK academic who had argued that, in part by repeatedly discussing a boycott of Israel, his trade union had created a hostile and anti-Semitic environment. Observers expect that the ruling will add fuel to this contentious fight.

Hawking, director of research at the Centre for Theoretical Cosmology, was scheduled to attend the Israeli Presidential Conference in Jerusalem, an event run in partnership with the Hebrew University of Jerusalem that will also celebrate the 90th birthday of Israeli President Shimon Peres. The gathering on 18–20 June will feature talks by scientists, other academics, artists and politicians on topics ranging from education and new media to political leadership.

Hawking wrote to the conference organizers on 3 May to say that he would not be attending so as to "respect the boycott" that some

researchers have called for over the past decade. His decision was announced last week on the website of the British Committee for the Universities of Palestine, a London-based group calling for academics to cut links to Israeli institutions, and triggered intense media attention.

The University of Cambridge initially said in a statement released last Wednesday that the physicist, who has the debilitating neurological condition motor-neuron disease, was withdrawing for health reasons. Later that day, it said that it had been mistaken and that Hawking's decision was "based on advice from Palestinian academics that he should respect the boycott".

Conference organizer Israel Maimon, a lawyer, thinks that Hawking's decision was wrong, and says that attempts to create an academic boycott of Israel are "outrageous and improper".

Such efforts stretch back a decade in Britain. In 2002, more than 100 academics signed a letter published in *The Guardian* newspaper calling for a moratorium on European funding of, and contracts with, Israeli institutions, in protest at the "violent repression against the Palestinian people in the occupied territories". Since then, various campaign groups and trade unions of teachers and lecturers have proposed boycott motions at their meetings, some of which have been approved.

Steven Rose, an emeritus professor of ▶

► biology at the Open University in Milton Keynes, UK, organized the 2002 moratorium call with his wife Hilary Rose, a sociologist and emeritus professor at the University of Bradford. He claims that in the past ten years the campaign has “gone on expanding”. Although there are no solid data on its effects, he says that moves by the Israeli government to counter it — such as 2011 legislation that outlaws boycott calls within Israel — are telling. “That shows it’s having an impact,” he says.

Other researchers question whether science in Israel has been significantly affected. The country collaborates with many institutions in Europe and the United States, and even with some academics in the occupied Palestinian territories. “I personally have not suffered from any boycott activity. I have relationships and collaborations with British scientists and scientists all over the world,” says Ruth Arnon, an immunologist at the Weizmann Institute of Science in Rehovot, Israel, and president of the Israel Academy of Sciences and Humanities in Jerusalem, who has strongly criticized Hawking’s decision and the boycott movement. “There may have been some who have been affected, but I don’t know of any real effect.”

Hawking’s announcement is a “huge development” that may aid the boycott campaign

in the United States, where it has traditionally not had much traction, says Jess Ghannam, a clinical psychiatrist at the University of California, San Francisco, and a member of the organizing committee of the US Campaign for the Academic and Cultural Boycott of Israel.

“There may have been some who have been affected, but I don’t know of any real effect.”

Supporters of a boycott draw parallels between their movement and the campaigns for an academic boycott of South Africa that occurred during that country’s apartheid regime. But critics say that Israel is being unfairly singled out, arguing that institutions and conferences in other nations with questionable human-rights records are not being boycotted. Some say that the focus on Israel is rooted in anti-Semitism, a charge that Rose and others strongly deny.

In another recent event, a UK employment tribunal ruled on 22 March in favour of the University and College Union (UCU), a trade union for higher-education teachers, in a case brought against it in 2011. Ronnie Fraser, a

“This is a real turning point, in terms of bringing attention to this issue.”

Supporters of a boycott draw parallels between their movement and the campaigns for an aca-

democratic teacher and a UCU member, had alleged “institutional anti-Semitism” that was in part due to repeated votes at the union’s annual meetings to boycott Israeli academic institutions. Although the union concluded in 2007, on the basis of legal advice, that a boycott would be unlawful, votes at subsequent meetings have approved related motions, such as discussing Palestinian calls for a boycott.

The tribunal dismissed Fraser’s complaint, calling it “an impermissible attempt to achieve a political end by litigious means”. Some academics fear that this could embolden the boycott movement. David Hirsh, a sociologist who researches anti-Semitism at Goldsmiths, University of London, and who helped to spearhead Britain’s anti-boycott Engage campaign, says that it gives a judicial stamp to the idea that such a boycott is not anti-Semitic. “That’s really damaging,” he says.

Meanwhile, Richard Axel, a neuroscientist at Columbia University in New York, says that he will go ahead with plans to attend the Jerusalem meeting next month. “I do have serious concerns over the Israeli occupation of Palestinian territories. I’m deeply troubled by it,” he says. “But I do not feel an academic boycott of Israeli universities will in any significant way make progress to a resolution of that crisis.” ■

RESEARCH INTEGRITY

Meeting targets lab lapses

Attendees search for ways to tackle misconduct and sloppy science.

BY RICHARD VAN NOORDEN

Every conference has its own brand of comedy, and the humour is deliciously dark when the subject is misconduct. James Kroll, who investigates misconduct allegations at the US National Science Foundation, knew he would get a laugh with his classic “excuses for plagiarism” slide, which included one scientist who blamed acid reflux, and another who was “distracted by bird vocalizations”.

But any ‘would-you-believe-that?’ jocular-ity at the 3rd World Conference on Research Integrity, held last week in Montreal, Canada, was light relief from the serious concerns that attendees had come to tackle. Allegations of misconduct are rising, retractions are on the up and concern is growing that sloppy lab practices are leading to unreliable research.

Experts debate whether the trends represent real increases or simply growing awareness. But attendees at the meeting were brimming with plans to combat problems ranging from out-and-out fraud to selective publication of

experiments. Among the potential solutions are spot audits of research data; independent replication of results; requirements for data-sharing; ethics codes and training; forced accountability for institutions; and greater protection for whistle-blowers. Still, the attendees acknowledged that it is hard to measure whether these strategies work — and hardest of all to provide incentives for change in a system in which scientists are rewarded for speedy success.

“We know that research misconduct is more common than expected,” Nicholas Steneck, an ethicist at the University of Michigan in Ann Arbor, told the meeting of more than 360 attendees from 45 countries, “but we don’t know if it is getting any worse and whether it can be prevented or deterred.”

In recent years, the number of misconduct-related inquiries has spiked at national oversight and advisory bodies (see ‘Questions of integrity’), and at research journals. Veronique Kiermer, executive editor of Nature Publishing

Group in New York, says that anonymous allegations are growing, “taking a toll” on editors who must investigate them, although many turn out to be ill-founded.

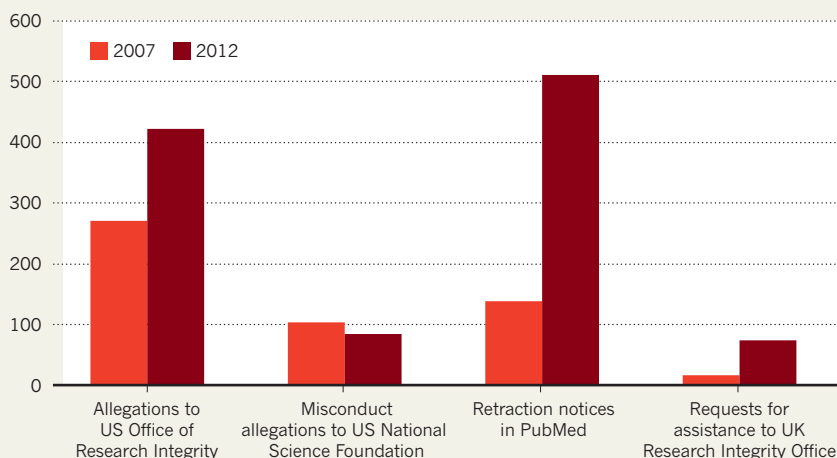
Lots of journals already scan manuscripts with plagiarism software. Bernd Pulverer, chief editor of *The EMBO Journal*, said at the meeting that his publication began systematic pre-publication examination of images in papers last November; around 4% contain “serious manipulations”, he said. But more discoveries do not mean that more misconduct is occurring. *The Journal of Cell Biology* has consistently revoked about 1% of accepted publications for image manipulation over the past decade, with no upward trend.

Meta-analyses suggest¹ that 1–2% of scientists admit to misconduct in anonymous surveys. A similar proportion admit to plagiarism, and 31% say that they have witnessed plagiarism by others, according to unpublished work by Daniele Fanelli, who studies science policy and misconduct at the University of Edinburgh, UK. “The field needs quantitative research — otherwise, we are just talking

► **NATURE.COM**
See Nature’s special
on reproducible
research:
go.nature.com/huhbyr

QUESTIONS OF INTEGRITY

In general, queries and allegations relating to misconduct increased between 2007 and 2012.



around our own assumptions,” he says. He thinks that analyses of the literature, for example to see when data look statistically improbable, will help to spot misconduct.

A new watchdog website, *Integru.org*, could also ferret out dishonesty. Launched last year to highlight evidence of plagiarism by scientists in Romania (see *Nature* **488**, 264–265; 2012), the site is now widening its focus so that anyone can submit cases in which they suspect plagiarism. The hope is that volunteer academics and the public will cooperate, using tools on the *Integru* platform to resolve each case.

For all the damage done by outright misconduct, meeting attendees said, lazy or sloppy research is a bigger concern. Institutions are now pouring resources into educating researchers about responsible conduct, but there is little evidence that the training is effective, according to the preliminary results of a systematic review presented at the meeting by Ana Marušić of the University of Split in Croatia.

Independent replication of studies — if it can be funded — could help to make sure that the literature is reliable. This idea has gained ground in the past year, after drug companies such as Bayer HealthCare in Leverkusen, Germany, and Amgen in Thousand Oaks, California, reported that they had trouble replicating published biomedical results². One study-validation mechanism, the Reproducibility Initiative based in Palo Alto, California, will task third parties with replication. This year, it has secured funding to replicate some 50 key biomedical experiments at a cost of around US\$20,000 each, said its founder, Elizabeth Iorns (see *Nature* **492**, 335–343; 2012). Almost 1,900 scientists have volunteered to have their results retested, she said.

Many ideas put forward at the conference — such as independent audits of raw data — are already in practice in clinical research, but not in bench science unless lab heads choose to enforce them. As far back as 1987, Adil

Shamoo, a biochemist at the University of Maryland in Baltimore, argued in *Nature* that scientists would have to expect routine data-auditing³. But the idea has never been accepted by the community, says Shamoo, who is also editor-in-chief of *Accountability in Research*.

One emerging concern is that modern research increasingly involves international collaborations. Melissa Anderson, who studies scientific integrity at the University of Minnesota in Minneapolis, says that integrity efforts will not be robust unless countries adopt common rules for factors such as ethical conduct and authorship. Simon Godecharle, a biomedical ethicist at the Catholic University of Leuven in Belgium, agrees. “Not one definition of research integrity or misconduct is the same in any two European countries” apart from Denmark and Norway, he says.

“Many countries are just at the start of their oversight integrity policies,” said Anderson. But compared with the first world meeting, six years ago in Lisbon, said Steneck, countries’ awareness that they need to foster integrity is at an all-time high. The difficulty now, he said, is in making sure that integrity codes and training requirements are actually affecting lab practice. “Everything we’re doing is above where the researchers live their day-to-day lives — and the challenge is to reach them.” ■

1. Fanelli, D. *PLoS ONE* **4**, e5738 (2009).
2. Begley, C. G. & Ellis, L. M. *Nature* **483**, 531–533 (2012).
3. Shamoo, A. & Annau, Z. *Nature* **327**, 550 (1987).

CORRECTION

The News story ‘Targeted drugs to tackle hepatitis C’ (*Nature* **497**, 18–19; 2013) wrongly implied that the US Preventive Services Task Force considers financial costs when evaluating the risks and benefits of screening. It does not.



PEDRO SAURA

Old masters

The earliest known cave paintings fuel arguments about whether Neanderthals were the mental equals of modern humans.

BY TIM APPENZELLER

In a damp Spanish cave, Alistair Pike applies a small grinder to the world's oldest known paintings. Every few minutes, the dentist-drill sound stops and Pike, an archaeologist from the University of Southampton, UK, stands aside so that a party of tourists can admire the simple artwork — hazy red disks, stencilled handprints, the outlines of bison — daubed on the cave wall tens of thousands of years ago. He hopes that the visitors won't notice the small scuff marks he has left.

In fact, Pike's grinder — and the scalpel that he wields to scrape off tiny samples — is doing no harm to the actual paintings, and he is working with the full approval of the Spanish authorities. Pike is after the crust of calcite that has built up over the millennia from groundwater dripping down the wall. The white flecks that he dislodges hold a smattering of uranium atoms, whose decay acts as a radioactive clock. A clock that has been ticking ever since the calcite formed on top of the art.

The results of an earlier round of sampling in El Castillo cave, published last June¹, showed that the oldest of the paintings, a simple red spot, dates to at least 40,800 years ago, roughly when the first modern humans reached western Europe. Pike and his colleagues think that when they analyse the latest samples, the paintings may turn out to be older still, perhaps by thousands of years — too old to have been made by modern humans. If so, the artists must have been Neanderthals, the

brawny, archaic people who were already living in Europe.

The answer won't be known for at least a year, but if it favours the Neanderthals, it could tip — if not resolve — a debate that has rumbled for decades: did the Neanderthals, once caricatured as brute cavemen, have minds like our own, capable of abstract thinking, symbolism and even art? It is one of the most haunting questions about the people who once shared a continent with us, then mysteriously vanished.

An early date for the paintings would also be a vindication for the slight, dark-haired man watching as Pike works: João Zilhão, who has emerged as the leading advocate for Neanderthals, relentlessly pressing the case that these ice-age Europeans were our cognitive equals. Zilhão, an archaeologist at the Catalan Institution for Research and Advanced Studies at the University of Barcelona in Spain, believes that other signs of sophisticated Neanderthal culture have already proved his point. But he is willing to debate on his opponents' terms. "To my mind, we don't need that evidence," he says of the paintings. "But I guess for many of my colleagues this would be the smoking gun."

The front line in the Neanderthal wars runs through another cave: Grotte du Renne, 1,000 kilometres away in central France. As early as the 1950s, excavations there unearthed a collection of puzzling artefacts. Among them were bone awls, distinctive stone blades and palaeolithic baubles — the teeth of animals such as foxes or marmots, grooved or

Spots and stencils in El Castillo cave, Spain — one at least 40,800 years old — might be the handiwork of Neanderthals.

pierced so that they could be worn on a string. They were buried beneath artefacts typical of the first modern humans in Europe, suggesting that these objects were older. A startling possibility loomed: that artefacts of this style, collectively known as the Châtelperronian industry, were made by Neanderthals.

Close cousins of modern humans, Neanderthals evolved in western Eurasia and had Europe to themselves for more than 200,000 years, enduring several ice ages. In spite of their survival skills and big brains — comparable to our own — they had never been linked to sophisticated tools of this kind, or to ornaments. Yet in 1980, archaeologists reported finding a Neanderthal skeleton among Châtelperronian tools at another site in France². And in 1996, French palaeoanthropologist Jean-Jacques Hublin and his colleagues reported that a skull fragment from the ornament layer in the Grotte du Renne was unmistakably Neanderthal³.

Ever since then, the Grotte du Renne has been exhibit A in the case that Neanderthals, like ourselves, trafficked in symbols, using ornaments as badges of identity for individuals or groups.

Hublin himself did not go that far. He suggested that the Neanderthals had fallen under the spell of strange new neighbours: modern humans, who were thought to have reached Europe around the time of the Châtelperronian industry. Neanderthals might have acquired the ice-age bling from modern humans, or made the pendants themselves under the influence of the new arrivals.

That conclusion infuriated Zilhão, turning him into the passionate advocate he is today. He questioned the evidence that modern humans were already on the scene and detected a bias against our extinct cousins. “Why was the equally if not more legitimate hypothesis — that the Neanderthals themselves had been the authors of this stuff and made it for their own use — not even considered?” asks Zilhão.

On a visit to rock-art sites in Portugal, he discussed the paper with Francesco d’Errico, an archaeologist who is now at the University of Bordeaux in France. D’Errico had the same reaction, Zilhão recalls. “And he said: ‘OK, let’s do something about it.’” Since then, the pair has fought a two-front war, advancing evidence for Neanderthal capabilities while challenging studies that reserve symbolism and abstract thinking for modern humans.

UNKNOWN ARTISTS

More than 15 years later, the Grotte du Renne continues to be a battleground. Since 2010, three papers have given duelling interpretations of the artefact-bearing layers. In the first, a group led by dating expert Thomas Higham of the University of Oxford, UK, used new carbon dates to argue that the layers were scrambled, mixing older remains with younger⁴. If that was correct, said Higham’s team, the relics adjacent to the telltale skull fragment might not have belonged to Neanderthals after all.

Within months, Zilhão, d’Errico and their colleagues fired back with an analysis⁵ of how artefacts of different types were distributed in the Grotte du Renne, concluding that the layers were undisturbed and that the Neanderthal link could be trusted. A group led by Hublin (now at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany) presented its own dates last year, backing Zilhão’s claim⁶. But Hublin still denied the Neanderthals full credit. Neanderthals did make the objects, now dated to between 45,000 and 40,000 years ago, he said — but only after they encountered modern humans. And this time he had fresh evidence to draw on.

Carbon dates measured by Higham and others at caves in Italy, Britain and Germany suggest that modern humans began expanding into Europe as early as 45,000 years ago, several thousand years earlier than was thought (see *Nature* **485**, 27–29; 2012). Zilhão strenuously disputes those claims, doubting whether the shells or animal bones used for dating truly reflect the age of the human fossils at the sites, or whether the human remains are modern. “The evidence to show an early presence of modern

humans in Europe is worse today than it was 20 years ago,” he declares.

Hublin, however, has no doubt that our ancestors had already entered the picture when Neanderthals in France began making bone awls and animal-tooth pendants. To assume that Neanderthals invented these technologies on their own is to accept “an incredible coincidence”, he says. “Just as modern humans arrive with these things in their pocket — bingo!”

LIKE MINDS

Despite the stalemate, Zilhão says that the record of Neanderthal behaviour tens of thousands of years before modern humans arrived in Europe proves his point (see ‘Minds at work’). Neanderthals are believed to have buried their dead, suggesting that they had some kind of spirituality. They made glue for securing spear points by heating birch sap while protecting it from the air, a feat that even modern experimental archaeologists have trouble replicating. Many Neanderthal sites include lumps of pigment — red ochre and black manganese — that sometimes seem to be worn down like stone-age crayons. Zilhão and others think that the Neanderthals painted themselves, creating striking patterns on their pale, northern skin that were every bit as symbolic as the art and ornaments of modern humans.

“You don’t need to have shell beads, you don’t need to have artefacts with graphical representation to have behaviour that can be defined archaeologically as symbolic,” he says. “Burying your dead is symbolic behaviour. Making sophisticated chemical compounds in order to haft your stone tools implies a capacity to think in abstract ways, a capacity to plan ahead, that’s fundamentally similar to ours.”

Where Zilhão sees a clear pattern, sceptics see uncertainties. Harold Dibble, an anthropologist at the University of Pennsylvania in Philadelphia, is re-examining supposed Neanderthal burial sites. At one, the French cave of Roc de Marsal, he says that what seemed to be a deliberately excavated grave is actually a natural pit. At another, La Ferrassie, he sees evidence that sediments swept into the cave by water — not grieving kin — could have buried Neanderthal remains.

As for the ochre crayons, Dibble is dismissive. “You see some wear on a piece of ochre and soon you’ve got Neanderthal body painting,” he says. “What a lot of logical leaps.” He and others say that the pigment has many possible uses: as an insect repellent, a preservative for food or animal skins, an ingredient in adhesives. Even Wil Roebroeks of the University of Leiden in the Netherlands, who found evidence for ochre use as early as 250,000 years ago at a Dutch Neanderthal site⁷, says that Zilhão “jumps too fast from the presence of ochre to body decoration”.

Ask Dibble, Hublin and other sceptics what would persuade them that Neanderthals had minds like ours, and their answer is simple: a pattern of art or other sophisticated symbolic expression from a time when no modern humans could possibly have been around. “But I don’t think it exists,” says Hublin.

Zilhão, however, points to a singular finding from a Neanderthal site in southern Spain that he reported three years ago⁸: three cockle shells each with holes near one edge, as if they had been worn as ornaments. One contains a trace of red pigment, and a fourth shell is stained with a mixture of colours, as if it had been used as a paint container. The shells, says Zilhão, imply symbolic thinking fully equivalent to that of the modern humans who left troves of beads in South Africa 75,000 years ago. And at roughly 50,000 years old, he says, the Spanish shells date from a time well before modern humans reached the region.

Critics are not satisfied. The perforations are natural, as Zilhão himself noted, which suggests to Hublin and Dibble that rather than systematically fashioning ornaments, Neanderthals might have picked up a few odd shells on a whim. “When you’ve got isolated occurrences, one-offs, that’s not going to convince most of us,” says Dibble.

The paintings in El Castillo could help to establish a pattern. The research group was conservative with the ages it reported last June¹, which put the earliest calcite at nearly 41,000 years old. Nervous about damaging the pigment, the team left several millimetres of the veneer intact at each

➔ NATURE.COM
To hear João Zilhão discuss Neanderthal minds, visit:
go.nature.com/zqnasp

MINDS AT WORK

Disputes over whether Neanderthals were capable of symbolic thinking turn on a smattering of discoveries spanning more than 200,000 years.

◆ Neanderthal ● Modern human ● Neanderthal or modern human

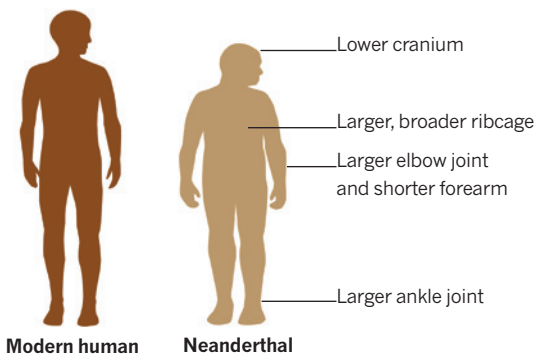
300,000 years ago (ya) First appearance of Neanderthals in Europe



Evidence suggests that not long after Neanderthals appeared in Europe, they were using ochre. The pigment can serve for decoration, but it also has practical uses.

250,000 ya Evidence of ochre use

Neanderthals were more robustly built than modern humans, which may have put different demands on their brains.



Modern human

Neanderthal



By heating birch sap while protecting it from air, Neanderthals made a glue for attaching stone flakes and points to spear handles.

120,000 ya Sophisticated hafting of spear points



Neanderthals are thought to have buried their dead, including this individual from La Ferrassie, France. But some researchers question whether the burials were deliberate.

70,000–50,000 ya Possible burials at La Ferrassie

50,000 ya Shell beads from Spain

Neanderthals may have worn pendants of coloured shells before the arrival of modern humans in Europe.

45,000 ya Modern humans appear in Italy

43,000 ya Modern humans appear in the United Kingdom

40,800 ya Oldest cave paintings in El Castillo, Spain

45,000–40,000 ya Animal-tooth pendants in France



sampled spot. Deeper, older layers might push back the paintings' minimum ages by several thousand years.

That prospect brought the team back to El Castillo last October. Grinding and scraping through a long day, the researchers concentrate on the red disks and hand stencils that had yielded the earliest dates last time around. The goal, says Zilhão, is "to date pigments in these paintings to an age that is clearly and to everyone's satisfaction beyond the range of modern humans in Europe".

Yet an early date may not settle the long-running dispute. Hublin sets the bar high. "If Zilhão finds a date of earlier than 50,000 years ago, I'll be convinced!" he says. Any younger, and modern human influence would remain a possibility, he says, noting recent hints that our ancestors had advanced into Turkey or even central Europe by 50,000 years ago. And one example of crude painting — what Dibble calls "Neanderthal doodling" — might not be enough to win over the doubters. Zilhão's knockout blow may simply lead to more fighting.

Yet signs of a middle ground are emerging. Chris Stringer, a palaeoanthropologist at the Natural History Museum in London, says that 20 years ago, he believed that if the Neanderthals made the Châtelperronian ornaments, they were blindly imitating modern humans. "Our interpretation was that they were copying but that they didn't have the brainpower to give full value" to the objects. He wouldn't say so now. Two decades of discoveries of sophisticated Neanderthal tools and weapons have made him think that "the gulf was not as great": that the difference between Neanderthals and ourselves was a matter more of culture than of ability.

"You can see the Neanderthals were held back by various factors that were not down to their brains," he adds. The climate of ice-age Europe kept their population size "frighteningly small", he says — at times just a few thousand people across a whole continent, most of them dead by the age of 30. How could such a sparse, beleaguered people develop and sustain a sophisticated culture?

That's not so different from what d'Errico, Zilhão's comrade-in-arms for almost 20 years, now says. He still thinks that the Neanderthals probably invented the Châtelperronian artefacts before modern humans were on the scene. But he is open to the idea that aspects of modern human culture preceded their wholesale arrival in Europe. "It's possible that some influence did spread," says d'Errico. "I'm less militant than João." That takes nothing away from the Neanderthals, he adds. "The fact that Neanderthals can absorb influences, can re-elaborate them, can make them part of their own culture, is very modern behaviour."

But there is a final stretch of ground that neither side will concede. Were the Neanderthals truly the same as us, cognitively? No, says Stringer. The Neanderthal genome, decoded⁹ in 2010, differs from that of modern humans in some regions linked to brain function, he notes. And this year, he suggested that, compared with modern humans, larger volumes of Neanderthals brains were devoted to vision and to controlling their heavier bodies¹⁰. That might have left them with less capacity for social awareness and interaction. "If you imagine a Neanderthal in modern society, there would still be differences," says Stringer.

Zilhão rejects any distinctions. Emerging from the cave into a rainy evening, he muses that if he pushes back the age of the El Castillo paintings, his critics may argue that he has simply proved an earlier presence of modern humans in Europe. "To which I will say, 'Of course. Neanderthals were modern humans too.'" ■

Tim Appenzeller is chief magazine editor at Nature.

1. Pike, A. W. G. *et al. Science* **336**, 1409–1413 (2012).
2. Lévêque, F. & Vandermeersch, B. *C. R. Acad. Sci. Paris* **291**, 187–189 (1980).
3. Hublin, J.-J., Spoor, F., Braun, M., Zonneveld, F. & Condemi, S. *Nature* **381**, 224–226 (1996).
4. Higham, T. *et al. Proc. Natl Acad. Sci. USA* **107**, 20234–20239 (2010).
5. Caron, F., d'Errico, F., Del moral, P., Santos, F. & Zilhão, J. *PLoS ONE* **6**, e21545 (2011).
6. Hublin, J.-J. *et al. Proc. Natl Acad. Sci. USA* **109**, 18743–18748 (2012).
7. Roebroeks, W. *et al. Proc. Natl Acad. Sci. USA* **109**, 1889–1894 (2012).
8. Zilhão, J. *et al. Proc. Natl Acad. Sci. USA* **107**, 1023–1028 (2010).
9. Green, R. E. *et al. Science* **328**, 710–722 (2010).
10. Pearce, E., Stringer, C. & Dunbar, R. I. M. *Proc. R. Soc. B* **280**, 20130168 (2013).

REF. 7

NATURAL HISTORY MUSEUM/MARY EVANS

ERIC LESSING/AGF-IMAGES

REF. 8



THE 18-KM² RAT TRAP

Ecuador has successfully eradicated invasive pigs and goats from most of the Galapagos archipelago. Now it is taking on the rats.

BY HENRY NICHOLLS

The helicopter appears as a speck on the horizon, moving slowly on a dead-straight path over the black volcanic island. Beneath it hangs a huge metal cone: an industrial-scale hopper that is sending a steady stream of blue pellets raining down on the scrubby landscape of Pinzón, one of the Galapagos Islands.

Erin Hagen watches through her binoculars. She is standing on the deck of the *Sierra Negra*, one of three vessels moored just off the island on this morning in November 2012. When the helicopter reaches the rocky shoreline, it changes course, heads across the ocean and hovers just above the boat. At Hagen's instruction, a team of conservationists comes to life. Two men stand by to service the hopper. Three others prepare to reload it with more than 400 kilograms of poisoned rat bait. Within three minutes, the loading is complete and the helicopter is heading back to lace Pinzón with more toxic bait. "It's kind of like a pit-stop," explains Hagen, a project manager for Island Conservation, an international non-governmental organization with expertise in the eradication of invasive species. "There's a real buzz from everyone."

Five years ago, most of the major islands and smaller rocky outcrops

in the Galapagos were home to a plague of invasive mice and rats. The rodents feed on the eggs and young of seabirds, land birds and reptiles, and have brought several species — including the rare Pinzón giant tortoise (*Chelonoidis duncanensis*) — to the brink of extinction. In 2007, the Galapagos National Park Service (GNP) and the Charles Darwin Foundation (CDF) developed an initiative code-named Project Pinzón, a military-style plan-of-action to kill invasive rodents on three islands — starting with North Seymour (1.8 square kilometres), then moving on to Rábida (5 square kilometres) and, finally, Pinzón (18 square kilometres) — plus around a dozen smaller outcrops and islets (see 'Rat race').

The effort, costing some US\$3 million so far, is not the biggest rat eradication ever attempted. But it is one of the most high-profile and challenging. Before conservationists and scientists could start attacking the rodents, they had to ensure that their poison would not take out some of the unique — and endangered — mockingbirds, finches, rails, iguanas and tortoises famously described by Charles Darwin. And whereas most rat eradications so far have targeted remote, uninhabited islands, the Galapagos is home to some 30,000 people and receives around 180,000 visitors each year. With so much boat traffic, the risk of reinvasion will be very high, says James Russell, an ecologist at the University of Auckland in New Zealand who has a special interest in rat invasions. "Their real challenge is going to be that biosecurity," he says.

For those involved, the anti-rat campaign is worth the trouble and the risks. It promises to allow unique species to flourish again and, building on the prior removal of feral pigs and goats from much of the archipelago, to make Ecuador a world leader in the eradication of invasive species. "Galapagos is up there in the front line looking to make the next big leap in multi-species pest management," Russell says.

By the time Darwin arrived in the Galapagos in 1835, the rodents had

CAROLYN JENKINS/ALAMY

long since settled in. Mice and black rats were probably the first to arrive, introduced by pirates or whalers in the seventeenth century; since the 1980s, Norway rats have found their way there too¹.

Galapagos experts have little doubt that the rodents have devastated native wildlife, even though the creatures' effects have not been studied systematically. "I just hated the immigrant killers because I could see what they were doing," says Felipe Cruz, a lifelong conservationist who grew up on Floreana, one of four inhabited islands in the archipelago. In the early 1980s, Cruz spent nine months of the year camped in the Floreana highlands deploying a cocktail of rodenticide to prevent rats from destroying eggs and chicks in the most important breeding colony of the Galapagos petrel (*Pterodroma phaeopygia*), a species that has been listed as critically endangered since 1994 (ref. 2).

His dedication was rewarded, says Cruz. "There were more birds, more plants, more lizards. It was like an island within an island." It was also a transformative experience. "It's something that fills me with pride, satisfaction and somehow shaped my life," he says.

SCALING UP

Before long, Cruz had an opportunity to eradicate black rats on a bigger scale: from the entire island of Pinzón, where they were destroying the island's endemic tortoise species by devouring hatchlings. In 1988, "there was a massive, massive drought and we began to see dead rats everywhere," recalls Linda Cayot, then at the Charles Darwin Research Station on the central island of Santa Cruz and now science adviser for the Galapagos Conservancy in Fairfax, Virginia. Cayot and Cruz saw the perfect opportunity to finish what the drought had started, and convinced the directors of the GNP and the CDF to allow a team to spread bait laced with rodenticide.

The rat population recovered after a few months, but Cayot describes the exercise as "a successful failure". The researchers realized that very young rats may not have encountered the bait, and that they needed to apply it twice. "I think we came really close," Cayot says. "We just learned so much about how to run a massive field operation."

This paved the way for an onslaught against much larger invasive mammals — mainly pigs and goats — whose relentless grazing had stripped out much of the vegetation, with knock-on consequences for native herbivores. Project Isabela, a staged initiative that began in 1997 and cost nearly \$10.5 million, resulted in the eradication of invasive pigs from the huge island of Santiago³ and some 140,000 goats from more than 5,000 square kilometres on several islands. According to Cruz and others involved in the project, it is "the world's largest island restoration effort to date"⁴. The effort also gave Galapagos conservationists the confidence to think big, says Cruz. They resolved to launch a new assault on the rats.

In 2007, the GNP and the CDF held a workshop to consider how best to approach the rat problem. This pulled together expertise from around the world, most notably from New Zealand, a country that has some 50 years of experience in the eradication of invasive species ranging from rabbits to wallabies. New Zealand is also the record holder for rat eradication: the 113-square-kilometre Campbell Island has officially been rat-free for almost a decade (<http://eradicationdb.fos.auckland.ac.nz>). "It's a bit of an industry in New Zealand," says John Parkes, a collaborator with Landcare Research, an environmental research organization based in Lincoln, New Zealand, and a key participant in the workshop. "There's lots of research that shows the benefits of removing these exotic species from islands far outweigh the short-term costs imposed by the control technologies themselves."

The workshop resulted in Project Pinzón, which, by working from small to bigger islands, had the aim of "taking on progressively larger

and more complex eradications", says Karl Campbell, senior programme director for Island Conservation. Soon after Ecuador's environment ministry signed off on the plan, the GNP and the CDF set to work on North Seymour, with Island Conservation entering into the project in 2008.

Later that year, with signs of success on North Seymour, Island Conservation approached Bell Laboratories, a company in Madison, Wisconsin, that specializes in industrial-scale rodent control. Would the firm donate sufficient bait — almost 45 tonnes — to cover all the remaining islets and islands identified in the Project Pinzón roadmap? The company, attracted by the philanthropic cause, agreed.

The biggest uncertainty was what impact the active ingredient — an anticoagulant called brodifacoum — might have on non-target species. In birds and mammals, the chemical prevents the repair of capillaries that rupture naturally, resulting in internal bleeding and, if the dose is high enough, death. What was not known was how the "eminently curious" fauna of the Galapagos, as Darwin described it, would respond if exposed to the bait. "We had to start building from scratch," Campbell says. This meant doing a risk assessment for all vertebrates and threatened species that might be affected on the islands yet to be baited.

Among the species of greatest concern were the Galapagos mockingbirds and finches, which might happily peck away at the bait. So, in 2009, graduate student Ana Lucía Carrión Bonilla of the University of San Francisco in Quito set about determining what colour these iconic birds like least⁵. Once she had the answer — blue — Bell Laboratories cooked up a blue version of an existing rat bait. Trials of a non-toxic version were promising: a suite of key endemic species turned their noses up at the bright blue blocks. Subsequent tests showed that a toxic version of the bait killed invasive house mice occupying the small uplifted island of North Plaza. In an added twist, the bait here and on most subsequent islands was spiked with a fluorescent dye to help researchers to track its movement through the environment by tell-tale traces on trails, nests, faeces and the animals themselves. "You go out with a big ultraviolet spotlight and you can see basically where this bait has gone," says Campbell. This suggested that some finches and lava lizards were nibbling at the bait, but were not killed by it.

Although the taste tests indicated that the Galapagos' world-famous tortoises and other reptiles would not gorge themselves on the bait, nobody was taking any chances. "We needed to get some hard data on what to expect if tortoises did eat the bait," says Penny Fisher of Landcare Research. In 2010, Fisher fed the bait pellets to captive tortoises of hybrid origin — which conservationists consider less valuable than

wild-living, purebred animals — then drew blood samples over the course of several weeks and measured how long they took to clot as an indicator of toxic effects. "It was a nerve-racking trial to do," she says. But, for reasons that are not yet clear, the coagulation time remained fairly constant, suggesting that tortoises would not face a serious risk of poisoning.

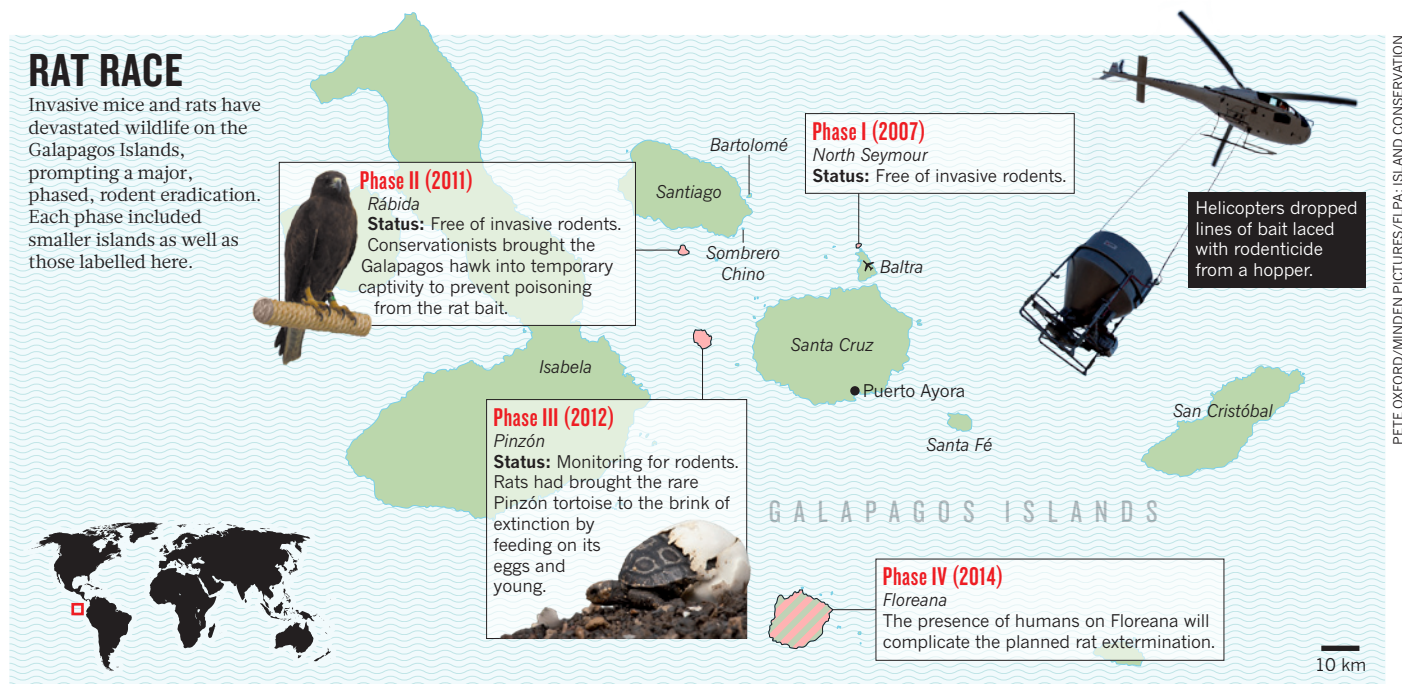
Protecting other species required more drastic measures. The Galapagos hawk perches at the top of the local food chain and, with a diet that ranges from young iguanas to sea lion afterbirth, faced the greatest risk of inadvertent poisoning. So the team decided to bring all territorial hawks into captivity until six weeks after the first bait drop. In January 2011, when the helicopter delivered bait over Rábida and a handful of smaller islets, Julia Ponder of the University of Minnesota looked after 20 hawks in makeshift aviaries on the nearby island of Santiago. All of the birds survived and were released back to the wild. And after a thorough survey in November 2012, the GNP declared Rábida free of invasive rodents.

Now, with a small population of tortoises in captivity as insurance, it was time to prepare for Pinzón itself. The helicopter pilot would need a three-day window of clear skies to bait the landscape with the necessary accuracy, and he would navigate along tight, pre-ordained flight

**"I JUST HATED THE
IMMIGRANT KILLERS
BECAUSE I COULD
SEE WHAT THEY
WERE DOING."**

RAT RACE

Invasive mice and rats have devastated wildlife on the Galapagos Islands, prompting a major, phased, rodent eradication. Each phase included smaller islands as well as those labelled here.



PETE OXFORD/MINDEN PICTURES/FLPA; ISLAND CONSERVATION

lines just 35 metres apart. “Keeping those lines straight is what makes or breaks a project,” says Hagen. “Then we know we can ensure full coverage.” In the run-up to the operation, Hagen and the baiting team practised loading and dropping bait on an abandoned US military airstrip on the island of Baltra. Then, in mid-November 2012 and with the weather forecast looking good, the operation began in earnest.

SIGNS OF SUCCESS

When Hagen stepped onto Pinzón after the second baiting in December, the island looked much as normal. “Typically what animals do when they are exposed to the rodenticide is go into safe areas because they don’t feel well,” she says. That means the only evidence of mass poisoning is the occasional smell of rotting flesh and perhaps a skeletal rat carcass. It usually takes around two years of monitoring — using traps for live animals, ‘bite cards’ distributed across the landscape, and searches for rat footprints and faeces — before an eradication effort is declared a success.

Assuming the rats are gone, the team will monitor how the ecosystem responds over the next five to ten years. Regular on-the-ground surveys will document how key species fare and acoustic sensors will gather data on the abundance and diversity of bird life from their calls. “For each of these threatened species the ultimate measure will be growing populations, possibly even self-sustaining,” says Nick Holmes, director of science at Island Conservation.

One archipelago-wide indicator will be land snails of the genus *Bulimulus*. Although not as famous as Darwin’s finches, these small snails with pine-cone-shaped shells offer an even more compelling illustration of natural selection’s creative force: there are some 70 documented species, all of which could be descended from a single common colonizing ancestor⁶. But rats feed on them, which might explain why more than 50 of the species are threatened — and their damaged shells offer a good way to quantify rat predation. “Invertebrates in general are going to give us a faster idea of the response to the eradication programme,” says Christine Parent of the University of California, Berkeley, who in years to come will help to monitor Pinzón’s snails for signs of a rebound.

But if there is one certainty in such a complex operation, it’s that not everything will go to plan. Although Ponder and her colleagues succeeded in keeping 60 Galapagos hawks in captivity over the course of the Pinzón campaign, the birds have not fared so well after their release. “We have 16 confirmed dead,” says Ponder. She thinks that the birds dined on small reptiles that had consumed the bait. In future, it

may pay to keep the hawks in captivity for longer.

Even if every last rodent is removed from an island, constant vigilance will be needed to prevent a reinvasion. In January 2011, the GNP and Island Conservation assessed the probability of reinvasion by baiting the islands of Bartolomé and Sombrero Chino, which lie within 500 metres — swimming distance for a rat — of the still rat-infested island of Santiago. In November 2012, they found evidence of rats on both of the treated islands. (Fortunately, Rábida and Pinzón are beyond rat swimming distance of neighbouring islands.)

Tourist vessels and other boat traffic will also need to be monitored for stowaway rats. And there is always the risk of human sabotage, which occurred several times after Project Isabela. In 2009, for instance, some malcontent set six goats down on Santiago, which by then had been goat-free for around three years. The GNP put the cost of monitoring the island and removing these animals at \$32,393, more than \$5,000 a beast⁴.

In spite of such setbacks, Ecuador’s environment ministry is pushing ahead with its programme of ecological restoration in Galapagos. For 2014, the central government has committed several million dollars to attempt the extermination of rats from the 173-square-kilometre island of Floreana, where the human population adds to the challenge. “Dropping poison around people adds a layer of complexity,” says Parkes. But success on an island the size of Floreana would set an example for the rest of the world. It would also make it easier to realize conservationists’ long-standing ambition of reintroducing mockingbirds and tortoises to the island, not to mention the respite it would bring for the Galapagos petrel.

And that would be the realization of a lifelong dream for Cruz. “When I travel between the islands and I get to see groups of petrels flying about, I have to be quite honest,” he says. “My heart beats faster.” ■

Henry Nicholls is a science journalist based in London. His latest book *The Galapagos* will be published by Profile Books in 2014.

1. Harper, G. A. & Carrion, V. in *Island Invasives: Eradication and Management* (eds Veitch, C. R., Clout, M. N. & Towns, D. R.) 63–66 (IUCN, 2011).
2. Cruz, J. B. & Cruz, F. *Biol. Conserv.* **42**, 303–311 (1987).
3. Cruz, F. et al. *Biol. Conserv.* **121**, 473–478 (2005).
4. Carrion, V. et al. *PLoS ONE* **6**, e18835 (2011).
5. Carrión Bonilla, A. L. *Preferencias de color de alimento en Pinzónes de Darwin y cucúes de Galápagos: Implicaciones para disminuir la muerte accidental por consumo de veneno*. Thesis, Univ. San Francisco de Quito (2009).
6. Parent, C. E. *Diversification on islands: Bulimulid land snails of Galapagos*. Thesis, Simon Fraser University (2008).

COMMENT

PHILANTHROPY A century of the Rockefeller Foundation, inventor of the grant **p.311**

BRAIN A history of HM, the most famous patient in neuroscience **p.313**

SPACE Buzz Aldrin's blueprint for colonizing Mars reviewed **p.314**



OBITUARY Remembering Robert Edwards, pioneer of *in vitro* fertilization **p.318**

PETE ELLIS/DRAWGOOD.COM



Don't bristle at blunders

Embrace mistakes, urges **Mario Livio** — they are portals to scientific progress.

In a July 1991 *Nature* paper¹, astronomers Andrew Lyne, Matthew Bailes and S. L. Shemar made an electrifying announcement: the discovery of the first planet outside our Solar System. To everyone's surprise, it was not orbiting a Sun-like star but a pulsar — the dense, spinning neutron-star offspring of a supernova explosion. The putative planet gave itself away by altering the period of radio-frequency flashes given off by the pulsar.

Unfortunately, Lyne and Bailes had to retract this result a few months later after uncovering an error, which they reported² in *Nature* in January 1992. The astronomers courageously announced that they

had not corrected adequately for Earth's motion around the Sun. Lyne's revelation of the blunder at a meeting of the American Astronomical Society that month won him a standing ovation. But the story had a happier ending.

Immediately after Lyne's presentation, astronomer Aleksander Wolszczan announced that he and his colleague Dale Frail had discovered two planets orbiting another pulsar using the same technique. These turned out to indeed be the first discoveries of extrasolar planets. Wolszczan told me that Lyne's original paper had acted as a "confidence booster", convincing him that the signals in his data were real. By the

time Lyne withdrew his result, Wolszczan had performed enough tests to be certain.

Blunders are an essential part of the scientific process. Research is not a linear march to the truth but a zigzag path, involving trial and error. Mistakes are not the exclusive province of sloppy or inexperienced scientists. Even the brightest luminaries — including Charles Darwin and Albert Einstein — made serious blunders.

Truly innovative ideas require a willingness to embrace risks, and acceptance of the fact that errors can be portals to progress. Although this is well known in some private companies engaged in research and development, academics today are slow ▶

► in recognizing the necessity of blunders. Chemist Linus Pauling knew it. His former postdoc, Jack Dunitz, recalls being told: “Mistakes do no harm in science because there are lots of smart people out there who will immediately spot a mistake and correct it. You can only make a fool of yourself and that does no harm, except to your pride. If it happens to be a good idea, however, and you don’t publish it, science may suffer a loss.”

KNOTTY PROBLEM

Preposterous ideas can lead to important insights. In 1867, the eminent physicist William Thomson (Lord Kelvin) proposed³ that atoms were not point-like but ‘knotted vortex tubes of the ether’. Ether was the supposed fluid that pervaded space, providing a medium for electricity and magnetism.

Inspired by work on vortices in fluids by the nineteenth-century German physicist Hermann von Helmholtz, Kelvin identified three characteristics of knotted vortex tubes that made them attractive models for atoms.

First, vortices in fluids were astonishingly stable — mirroring to Kelvin the “unalterable distinguishing qualities” of atoms — and each knot could be classed according to its geometrical properties. Second, the variety of chemical elements could reflect the “endless variety” of knots. Finally, just as smoke rings vibrate, the oscillations of ether vortex tubes might produce atomic spectral lines.

To explain the periodic table, Kelvin needed to classify knots according to their forms, discarding any that could be manipulated from one to another. In Kelvin’s theory, the circular ‘unknot’ represented the hydrogen atom; the triple-looped ‘trefoil’, carbon.

Kelvin’s theory of vortex atoms is obviously wrong. The ether does not even exist. But these failures did not deter everyone. Whereas physicists lost interest for a while, knots began to intrigue mathematicians, becoming an active area of study for decades.

In the 1980s, knot theory reconnected with physics. Mathematician Vaughan Jones discovered an algebraic expression that is unique for every knot. Physicist Edward Witten linked it to quantum field theory, the branch of physics that describes fields and the subatomic world. In classical physics, the path of a particle travelling from point A to point B is determined by Newton’s laws of motion. In the quantum regime, one has to consider all the possible paths connecting A to B, including winding and knotted ways.

Subsequent work linked knots, quantum field theory and string theory, which by describing particles as vibrations of strings, harks back to Kelvin’s idea. Today, knots are used in chemistry and biology to analyse the actions of enzymes on DNA molecules. In a process known as site-specific recombination, enzymes align segments of the genetic sequence, cut the two strands of DNA open

and recombine the four ends in various ways, which can be described using knot theory.

EXTRAORDINARY CLAIMS

Blunders are sometimes hard to correct. Modern experiments can be so intricate and require such big investments in time and funds that replicating them becomes prohibitive. When a result is widely assumed to be wrong, few scientists are motivated to repeat the work.

But there can be rewards for doing so. The sensational claim⁴ in *Science* by geomicrobiologist Felisa Wolfe-Simon and her colleagues to have discovered a bacterium that substitutes arsenic for phosphorus to sustain its growth brought a wave of criticism.

A few critics checked the experiment, including microbiologist Rosemary Redfield at the University of British Columbia in Vancouver, Canada, who blogged the process (see go.nature.com/bmb62d). The effort proved fruitful, showing that the bacterium goes to great lengths to dodge arsenic. Redfield and her colleagues detected no arsenic in the bacterium’s DNA to much lower limits than in the original paper. Molecular biologist Dan Tawfik and his team at the Weizmann Institute of Science in Rehovot, Israel, identified the mechanism by which some of the proteins of this and related bacteria bind to phosphate and not to arsenate.

Although one lesson is obvious — extraordinary claims require extraordinary evidence — the original paper still had some scientific value. It stimulated discussion and inspired curiosity about different possibilities for life.

In the nineteenth century, Scottish author Samuel Smiles wrote: “We often discover what will do, by finding out what will not do; and probably he who never made a mistake never made a discovery.” His statement should not be taken as advocacy for slapdash science but as an encouragement to think originally and take calculated risks.

Can research failure be accommodated in today’s fast-paced, funding-starved,

HISTORY REVISITED

Did Einstein ever say “biggest blunder”?

Almost any history of Albert Einstein’s ‘cosmological constant’ mentions his “biggest blunder” — the introduction of this constant to counteract gravity into equations characterizing the Universe.

Did Einstein actually say this? After scrutinizing dozens of documents while researching my book *Brilliant Blunders* (Simon & Schuster, 2013), I found no evidence that he did.

The “biggest blunder” phrase seems to have come from the colourful physicist

George Gamow in an article published in the September 1956 issue of *Scientific American*. Gamow later repeated the story in his 1970 autobiography, *My World Line*.

Einstein was indeed unhappy about having introduced the cosmological constant, saying in a letter to cosmologist Georges Lemaître that he was “unable to believe that such an ugly thing should be realized in nature”. Calling it the “biggest blunder” was, in my view, Gamow’s hyperbole.

impact-driven atmosphere? I believe it must. We should make space for risky scientific proposals in grant and evaluation processes.

Until a decade ago, the committees that allocated observing time on the Hubble Space Telescope were encouraged to give up to 10% of the time to proposals with a low probability of success but potentially high return. A similar philosophy could be adopted more widely.

One problem is that committees tend not to approve risky programmes. Efforts to reach consensus converge to a mean. Such obstacles can be overcome if decisions are left to one person. In the case of Hubble, a pool of ‘director’s discretionary time’ on the telescope is available, for which anyone can apply. From it came the Hubble Deep Field, one of the most detailed images of the Universe ever made.

Today, telescopes including Hubble are turned towards addressing the profound outcome of another ‘blunder’. Einstein regretted his attempt to model a static cosmos using a repulsive-gravity force (see ‘Did Einstein ever say “biggest blunder”?’). But since it was revealed by supernovae observations in 1998 that our Universe is accelerating, understanding the nature of that repulsive force is one of the biggest challenges that physics faces today.

Researchers must embrace blunders that come from thinking outside the box. Evaluation processes should allow for originality, even at the risk of false starts and blind alleys. ■

Mario Livio is an astrophysicist at the Space Telescope Science Institute, Baltimore, Maryland, USA. His book *Brilliant Blunders* (Simon & Schuster) is released this month. e-mail: mlivio@stsci.edu

1. Bailes, M., Lyne, A. G. & Shemar S. L. *Nature* **352**, 311–313 (1991).
2. Lyne, A. G. & Bailes, M. *Nature* **355**, 213 (1992).
3. Thomson, W. *Proc. R. Soc. Edinb.* **6**, 94–105 (1867).
4. Wolfe-Simon, F. et al. *Science* **332**, 1163–1166 (2011).



Vaccinating against yellow fever in Brazil.

The difficult art of giving

William H. Schneider reflects on the centenary of the Rockefeller Foundation, which began the postdoc and the grant, and led to the World Health Organization.

The board of John D. Rockefeller's eponymous foundation met for the first time 100 years ago this month, in May 1913. The size and mission of the Rockefeller Foundation were unprecedented. With its initial pot of US\$100 million (equal to \$2.3 billion today), it aimed to "promote the well-being of mankind throughout the world".

Rockefeller focused his philanthropy on health, following the recommendation of his first adviser, Frederick T. Gates. "Disease," said Gates, "is the main source of almost all other human ills, poverty, crime, ignorance, vice, inefficiency, hereditary taint, and many other evils."

The influence of the Rockefeller Foundation on the history of science and medicine has been profound. It has supported research by leading international scientists, helped to create and modernize the world's major medical schools and established the infrastructure and model of global-health programmes to combat disease. In the first half of the twentieth century, the foundation set the standard for big philanthropy, invented the grant proposal and implemented a worldwide system of fellowships

to train junior researchers.

Although the Rockefeller Foundation is now eclipsed in scale by a growing number of public and private major funders, its legacy continues to be felt. Today, it focuses on transforming cities, revaluing ecosystems, securing livelihoods and advancing health.

INSPIRING MEN

Two things helped the foundation to dominate scientific research for its first four decades: the changes wrought by the First World War and an innovative group of foundation officers. The war decimated European dominance in science, offering an opening for the Rockefeller Foundation to rebuild and develop medicine and public health internationally as the US government pursued a policy of isolationism.

Of the coming of peace in 1918, John D. Rockefeller Jr, son of the founder and first president of the foundation, wrote, "In this commerce of culture, science, sympathy, and idealism, the Rockefeller Foundation desires to put its policies, personnel, and resources at the service of the world."

The height of the foundation's influence

was in the 1920s and 1930s. After making a 5-year, \$500,000 grant to the US National Research Council in 1919 for postdoctoral fellowships in chemistry and physics, the Rockefeller Foundation began funding its own fellowships for scientists to train at established centres worldwide.

The rationale and strategy for the fellowships were proposed in 1916 by Wickliffe Rose, who headed the foundation's International Health Board. Rose found that there was a shortage of good people working in international health. His solution was to "locate the inspiring productive men" and provide fellowships "for the international migration of select students" to train with them — as long as the students agreed to return to their own countries afterwards. He also proposed giving some assistance "in the more backward countries, to enable the returned scientists to establish themselves".

Between 1917 and 1950, the Rockefeller Foundation funded 6,342 fellowships. Almost half were in medical education and the sciences. Nobel-prizewinning physicist Robert Millikan, who was heavily involved, wrote in his 1950 autobiography that the fellowships were "the most effective agency in the scientific development of American life and civilization ... in my lifetime". To identify mentors, foundation officers and field staff surveyed major schools and research labs around the world. Reports of medical education alone were written for 30 European countries between 1920 and 1931.

Another main goal was reform of medical education. Beneficiaries included Johns Hopkins University in Baltimore, Maryland; Harvard University in Cambridge, Massachusetts, and the University of Toronto in Canada. Among the European recipients were University College London and the medical school at the University of Brussels. The Peking Union Medical College in Beijing was lavishly funded (a total of \$45 million was given by 1947) with the explicit goal of creating the 'Johns Hopkins of Asia'. The foundation also gave major grants to medical research institutes in the United States (such as Yale University in New Haven, Connecticut), Canada (to McGill's Montreal Neurological Institute) and Europe (to the Institute of Physiology in Copenhagen and the Tavistock Clinic in London). Funding of the Max Planck Institute of Psychiatry in Munich, Germany, and the Kaiser Wilhelm Institute of Anthropology, Human Heredity, and Eugenics in Berlin, which conducted eugenic research during the rise of Nazism in the 1930s, brought opprobrium at the time and since¹.

A third major area of support was tens of millions of dollars in grants to international and public health, notably to establish schools at Johns Hopkins (designed to be the model for US universities), Harvard and

the London School of Hygiene and Tropical Medicine. The foundation gave lesser grants to establish public-health laboratories in the newly created countries of eastern Europe² and at the Oswaldo Cruz Institute in Rio de Janeiro, Brazil.

The resulting global network of public-health institutions paved the way for the creation in 1948 of the World Health Organization (WHO), a United Nations specialized agency. The WHO built on the worldwide Rockefeller Foundation campaigns against diseases such as hookworm, yellow fever and malaria, leading to the launch of the WHO Global Malaria Eradication Programme in 1955, a precursor to its landmark smallpox eradication campaign, which began a few years later.

MAKING THE PEAKS HIGHER

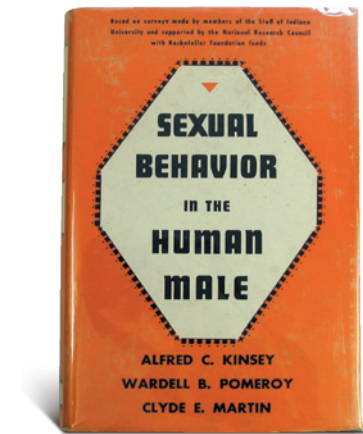
Another of the Rockefeller Foundation's key innovations was the research grant. After the financial crash that led to the Great Depression, a 1929 plan of action by the trustees recommended that "the development of medical schools as institutions be lessened and the principle of aid to individuals, groups, and departments in relation to research and advance of medical knowledge be emphasized".

The result was the use of short-term, targeted assistance for new fields such as molecular biology³. Biologist Alfred Kinsey was given a grant for research on human sexuality, which a governmental agency would have been unlikely to fund. The foundation also helped to establish psychiatry as a scientific discipline at major US medical schools such as Yale, Chicago, Harvard and Washington University in St. Louis, Missouri⁴.

To identify the best researchers and evaluate projects, the Rockefeller Foundation relied primarily on its small cadre of long-serving programme officers. They pursued a strategy of 'making the peaks higher', funding people and programmes that were already strong.

This policy signalled the beginning of the grant model of research support, which now dominates science. And although it increased the foundation's flexibility, at least one programme officer presciently recognized its disadvantages.

"We are really in many cases a university playing hide and seek in other universities' buildings," warned Alan Gregg, head of the Rockefeller Medical Sciences Division, in 1937. Gregg, who worked at the foundation from 1919 to 1956, went on: "We select the kind of work to be done (or they sagaciously anticipate our wishes), we choose the personnel and we pay the bill. But we don't accept the responsibility for personnel or for the future." The result, he feared, was "fragmentary and inconclusive" work, "timid



Work funded by an early Rockefeller grant.

and restless" employees, and "uncertain or resentful" universities⁴.

If the First World War fostered the rise of the Rockefeller Foundation, the Second World War did the opposite. After 1945, the US government, following unprecedented investment in science during the war, overtook the Rockefeller Foundation as the major funder of such work. It greatly expanded funds for external research by the National Institutes of Health (NIH) and created a new National Science Foundation (NSF). Grants given in 1953 by the Rockefeller medical and natural sciences divisions amounted to around \$9.2 million; that year, the NIH gave more than \$22 million to extramural medical research. By 1959, the NIH figure reached \$156 million⁵, and the NSF gave \$49 million to basic science research.

In 1949, Gregg reported that the funding policies of the NIH had been derived "more from Rockefeller experience than anywhere else". Ever the critic of small fixed-term grants, he quipped, "the Foundation no longer risks becoming the largest distributor of chicken feed in the country"⁶.

There were noteworthy differences between the Rockefeller and the NIH. To obtain congressional funding, the NIH set out to increase the geographical distribution of grantees. And it relied on peer review, with proposals evaluated by standing committees of experts, divided by research specialities. As early as 1946, there were ten study sections of the NIH that reviewed proposals in fields such as antibiotics, biochemistry, pathology, malaria and syphilis⁷.

At the same time, new foundations were established that equalled and overshadowed the influence of the Rockefeller in private philanthropy, forcing a change in strategy. The Ford Foundation, established in 1936 by car magnate Henry Ford and his son Edsel, gave \$300 million in grants in 1954, and almost double that (\$570 million) the following year⁸. The Rockefeller shifted to funding research in agricultural production and population control, the humanities and arts, and new fields such as international

relations, civil rights and the environment⁹.

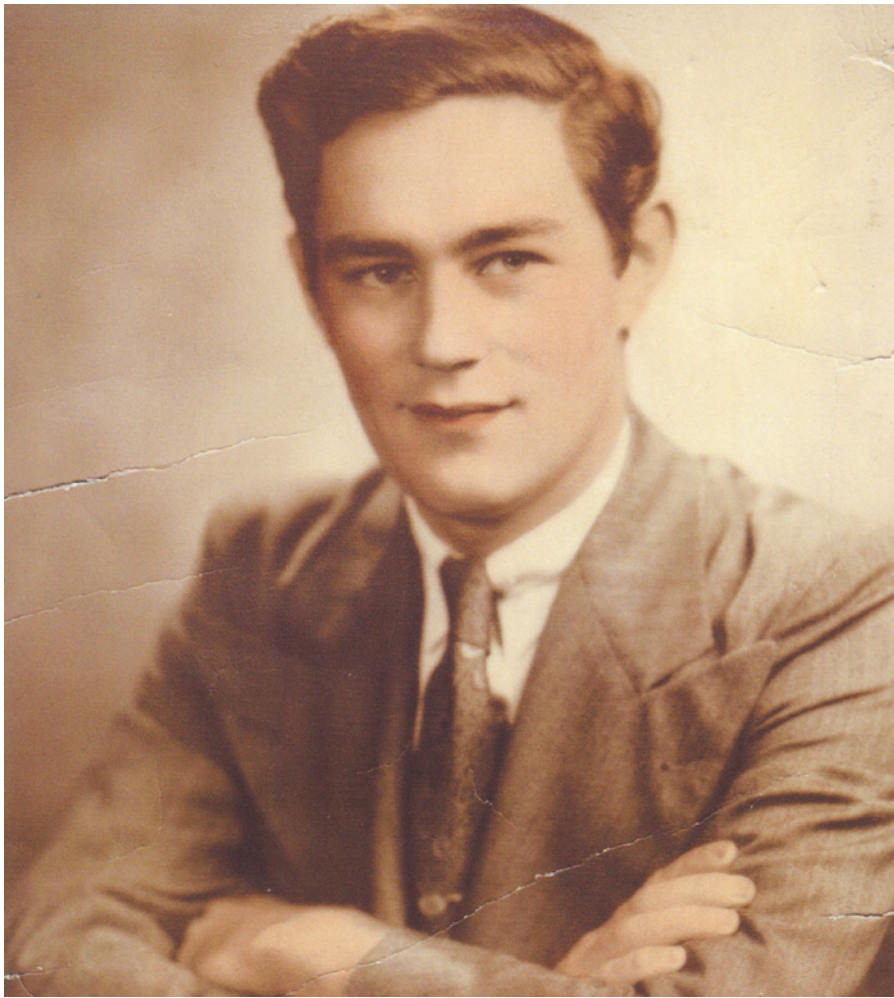
Today, a huge philanthropic entity again dominates the funding scene and focuses on health — the Bill & Melinda Gates Foundation, headquartered in Seattle, Washington. It too was established by a man who had made his fortune in business. The Gates Foundation 2011 annual report listed almost \$35 billion in assets, more than three times those of the next biggest US foundation (Ford) and ten times those of the Rockefeller, now the 16th largest. Bill and Melinda Gates' decision to focus on health in poor countries, like Rockefeller's a century ago, shows compassion and is a way to counter criticisms of the business activities that generate wealth. But the Rockefeller spent even more on fundamental research and programmes to reform medicine in advanced countries.

The Gates relies more heavily than the Rockefeller did in its heyday on the academic and scientific community for guidance and review of its funding strategies and activities, in effect, following the NIH model. And it is more apt to partner with other foundations and private businesses, something that the Rockefeller Foundation began to do in only the 1950s. Both foundations have been criticized for an overemphasis on technology to solve problems that have significant political, social and economic roots¹⁰.

When the Gates Foundation celebrates its 100th anniversary in 2094, historians will be better able to compare the successes of these major philanthropic organizations. For now, the Rockefeller legacy bears out what industrialist Andrew Carnegie predicted in 1889 at the dawn of the age of philanthropic foundations: "It is more difficult to give money away intelligently than to earn it in the first place." ■

William H. Schneider is editor of *Rockefeller Philanthropy and Modern Biomedicine* (Indiana University Press, 2002). He is professor of history and of *philanthropic studies* at Indiana University, Indianapolis, USA.
e-mail: whschnei@iupui.edu

1. Richardson, M. *Minerva* **28**, 21–58 (1990).
2. Weindling, P. *Minerva* **31**, 253–267 (1993).
3. Kohler, R. F. *Minerva* **14**, 279–306 (1976).
4. Schneider, W. H. *Rockefeller Philanthropy and Modern Biomedicine: International Initiatives from World War I to the Cold War* (ed. Schneider, W. H.) 7–60 (Indiana Univ. Press, 2002).
5. US National Institutes of Health. *Basic Data Relating to the National Institutes of Health 1966* 22 (NIH, 1966).
6. Penfield, W. *The Difficult Art of Giving: The Epic of Alan Gregg* 295 (Little, Brown, 1967).
7. Mandel, R. *A Half Century of Peer Review, 1946–1996* (Division of Research Grants, NIH, 1996).
8. Nevins, A. *Ford: Decline and Rebirth, 1933–1962*, vol. 3, 417 (Scribner's, 1963).
9. Rockefeller Foundation. *Annual Report, 1953* 36–43 (Rockefeller Foundation, 1953).
10. Birn, A.-E. *Lancet* **366**, 514–519 (2005).



Henry Molaison on his graduation day.

NEUROSCIENCE

Losing the past

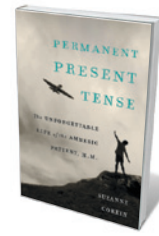
Douwe Draaisma visits the unusual mind of Henry Molaison, the most famous patient in brain science.

On 2 December 2008, Henry Molaison died. He was 82 and had been living in a nursing home in Windsor Locks, Connecticut. After Suzanne Corkin, a neuroscientist at the Massachusetts Institute of Technology in Cambridge, issued a press release about his death, it became clear that Molaison was the elusive 'HM' — arguably the most famous patient in brain science.

HM's fame hinged on a calamity. His first life had ended long before, in 1953, when a surgeon's knife left a devastating loss of memory in its wake. The nature of his lesion and its consequences launched HM on a distinguished career as a subject of many key neuropsychological experiments. Corkin, who worked with HM for half a century,

has now written *Permanent Present Tense*. She has woven her memories of her experimental and personal dealings with HM into a panoramic history of the past 60 years of the neuropsychology of memory. The result is superb. Because she was such an integral part of this history, *Permanent Present Tense* is also her intellectual autobiography.

Molaison was just 27 when he underwent the devastating surgery to treat his intractable epilepsy. His operation was "frankly experimental", as his surgeon, William Scoville, put it. The amygdala and a large part of the hippocampus were removed from both sides of Molaison's brain. Textbook versions of HM's story often feature 'a daring surgeon', but some feel that a bilateral resection



Permanent Present Tense: The Unforgettable Life of the Amnesic Patient, H.M.

SUZANNE CORKIN
Basic Books: 2013.
400 pp. \$28.99

was an irresponsible act, even if it was common practice in the rough field of lobotomy. The surgery brought Molaison partial relief from seizures, but expunged most of his memories and left him unable to store new information. From then on, HM was frozen in the titular 'permanent present tense'.

Corkin first met HM in 1962, and studied him for her PhD thesis in Brenda Milner's lab at the Montreal Neurological Institute in Canada. After Corkin's long collaboration with HM, she was still only a vaguely familiar face to him. She acted both as a researcher and a gatekeeper, carefully scrutinizing fellow scientists and their research plans before allowing access to this precious test subject. Unlike Alzheimer's disease, which causes global neurological damage, HM's lesion was rather specific. And so, it seemed, was the nature of his amnesia. Yet some of the momentous episodes in his 'career' as a study subject were associated with findings that complicated the initial image of simple anterograde amnesia.

To begin with, work with HM offered experimental proof of the distinction between short-term and long-term memory. HM could store information — such as a sentence — for about 10 or 20 seconds. Once he stopped thinking about it, there was no way it could re-enter his short-term memory. Apparently, the hippocampus is essential for moving information from short-term to long-term memory, but not the other way round. Some of his spared older memories could pop up in his mind and prompt him to tell an anecdote from his youth, even if he was unable to remember that he had told the same story 10 minutes ago.

In 1962, however, Milner demonstrated with HM that some types of information did make their way into long-term memory. After three daily sessions during a week's training on a visuomotor task — tracing a star shape while looking in a mirror — HM's performance improved considerably. He forgot both lessons and teacher, but remembered what he had learned. Preserved learning was confirmed in other types of tasks, and contributed much to the distinction between

nondeclarative (or implicit) and declarative memory. To most of us, knowing that one has learned something and remembering ▶

➤ **NATURE.COM**
A retrospective look
at Jorge Luis Borges
and memory:
go.nature.com/5de4vm

▶ what one has learned seem to be the same thing. HM showed that these are in fact two separate types of memory.

Further work with HM also forced Corkin to reconsider the initial idea that his operation had effectively cut his memory into two neatly separated temporal compartments. This was true for his vocabulary: his world remained one of rocketeers, never of astronauts. But when Corkin asked HM in 1966 to draw the floor plan of the house he had moved into five years after his operation, he produced an accurate plan from memory. When he moved house, however, HM would get lost and show up at the doorstep of his former home.

Several details on HM's personal life were already available in 1995, when science writer Philip J. Hilts published his compelling book *Memory's Ghost* (Touchstone, 1996). Where Hilts and Corkin seem to diverge is in their appreciation of Scoville's operation. Corkin does sketch Scoville's questionable background in lobotomy, but is somewhat apologetic on the bilateral resection. Hilts is considerably more critical, mentioning that Montreal neurosurgeon Wilder Penfield "exploded into the telephone" when Scoville called him with the details of HM's surgery. Ironically, the Scoville and Milner 1957 paper, 'Loss of Recent Memory after Bilateral Hippocampal Lesions', became a citation classic and the operation on HM developed into Scoville's claim to fame.

In 2002, Corkin announced that HM had agreed to donate his brain to science. She decided that Jacopo Annese, head of the Brain Observatory at the University of California, San Diego, was the person for the job. HM's brain even raised funding after his death: for US\$50, visitors to the Brain Observatory site could sponsor one of the 2,401 slide-mounted sections into which it had been sliced.

The exact locations of the lesions that caused HM's amnesia were ascertained by high-resolution magnetic resonance imaging scans, made a few hours after his death. HM's virtual brain will probably take its place among relics of neurology such as the skull and iron rod of Phineas Gage and the brain of the speechless Monsieur Tan, neurologist Paul Broca's patient. Whether it will serve as an opportunity for future research remains to be seen, but HM will live on in other ways. The rights to make a film about his life have been sold to Columbia Pictures and Scott Rudin, who in 2007 produced *No Country for Old Men*. ■

Douwe Draaisma is professor of the history of psychology at the University of Groningen, the Netherlands. His book *The Nostalgia Factory* will be published in September 2013. e-mail: d.draaisma@rug.nl



An artist's impression of a future mission to the red planet.

SPACE EXPLORATION

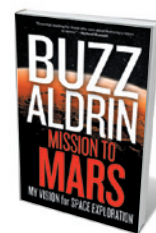
Life on Mars

Jim Bell welcomes a detailed blueprint for colonizing the red planet from *Apollo 11* veteran Buzz Aldrin.

For many, our exploration of the cosmos seems to hinge on grand and singular theoretical observations and technological achievements such as those of Galileo and Einstein. We are fascinated by paradigm-shifting discoveries and personal stories — the oppressed astronomer, the quirky professor. Yet in the arena of global science and engineering, it often takes the less flashy, more mundane development of good policy and follow-on infrastructure to maintain and sustain those changes, and to enable even more discoveries.

In *Mission to Mars*, celestial-mechanics expert and *Apollo 11* astronaut Edwin ('Buzz') Aldrin, with space journalist Leonard David, argues that achieving the next great singular goal in space exploration — a sustained campaign to explore and ultimately settle Mars — will require a healthy dose of such pragmatism. The United States will, they say, need both significant advances in its space transport infrastructure and considerable changes in space policy.

Ever since his historic 1969 lunar-landing mission, Aldrin has been a passionate and outspoken advocate for continued US leadership in space exploration. Calling the space shuttle programme "bad judgment" because it placed humans and cargo together, and a second race to the Moon "a



Mission to Mars:
My Vision for
Space Exploration
BUZZ ALDRIN
National Geographic
Society: 2013. 272 pp.
\$26 (£17.99)

dead end" in terms of either a commercial or a scientific pay-off, here Aldrin plants his flag in the middle of NASA's ongoing identity crisis about the future of crewed space missions. He draws a line in the exosphere, defining low-Earth orbit and even the Moon as the realm of private enterprise and commercial space development. He goes on to write that "America's space program should help other nations achieve what we have already done", and echoes the controversial call to bring both China and India in as International Space Station partners.

These and other recommendations from one of America's first moonwalkers make *Mission to Mars* more than just a plea for exploration of the red planet. Aldrin outlines what he calls his "Unified Space Vision", which he believes is needed to enable humanity to establish permanent settlements beyond Earth. Some of its crucial elements would be an independent think tank to advise government, industry and the public on space issues,

NASA/PAT RAWLINGS, SAIC

an international lunar-development authority to oversee the eventual economic exploitation of the Moon, and a planetary-defence component to advance our understanding of the science and mitigation strategies for potentially hazardous near-Earth asteroids.

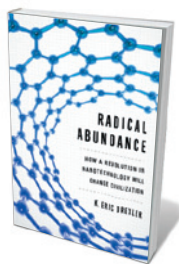
Perhaps Aldrin's most passionate call, however, is for the development of a fully reusable interplanetary transport infrastructure between Earth and Mars. This would break through the current 'reusability barricade' for space travel, an achievement that was visualized for the space shuttle, but never fully realized. A key advance would be to implement Aldrin's long-standing idea of 'cyclers' spacecraft: ferries set in motion on permanent looping trajectories around the gravitational orbits of Earth and Mars (or other staging destinations, such as the Martian moons). There are many technical issues that limit such a concept: for example, passengers to and from these destinations would have to use some sort of shuttlecraft to catch up with the non-stop, high-velocity cyclers. But Aldrin rightly points out that without the infrastructure in place for transportation, as well as for essential long-term resources such as fuel, food and water, we will continue to make only small steps, rather than the next giant leaps, in our exploration of the worlds around us.

I was perhaps most struck by Aldrin's rejection of the idea of returning to the Moon before any foray to Mars. "Don't put any more NASA astronauts on the Moon!" he implores. Indeed, his philosophy is consistent with the current space policy of President Barack Obama's administration, which includes NASA's recently announced idea to attempt to capture a small near-Earth asteroid. This mission is poorly defined at present, but if eventually implemented it could be consistent with Aldrin's call for testing new deep-space rocket components, and for potential scientific and planetary-defence research on near-Earth asteroids. Some of Aldrin's ideas may be having an impact at the highest levels.

We idolize aviation pioneers and their many 'firsts', but take for granted today's global air-transport infrastructure predicated on their achievements. Even the Global Positioning System can be traced back directly to the ventures of NASA and other space agencies. Buzz Aldrin doesn't want us to forget earlier heroic achievements. But he desperately wants us to get to work on the infrastructure and policies needed to make humanity a multi-planet species with an interplanetary economy. ■

Jim Bell is a planetary scientist and professor in the School of Earth and Space Exploration at Arizona State University in Tempe. He is president of the Planetary Society, the world's largest public space-advocacy organization. His latest book is *The Space Book*. e-mail: jim.bell@asu.edu

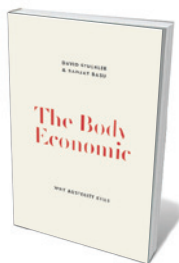
Books in brief



Radical Abundance: How a Revolution in Nanotechnology Will Change Civilization

K. Eric Drexler PUBLICAFFAIRS 368 pp. \$28.99 (2013)

Nanotechnology pioneer Eric Drexler bids us to leap in at the technological deep end. We can transform the way we make everything from bridges to circuit boards, he argues, by harnessing molecular machines that operate on digital principles. The result? Desktop or garage facilities that use less fuel, land and energy than today's vast factories and supply chains. The technical and political challenges of unleashing 'atomically precise manufacturing' are substantial, but Drexler cuts deftly through the complexities.



The Body Economic: Why Austerity Kills

David Stuckler and Sanjay Basu BASIC BOOKS 240 pp. \$26.99 (2013)

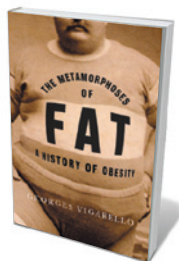
What price a healthy stock market? In this stringent economic analysis, sociologist David Stuckler and epidemiologist Sanjay Basu argue that during a recession, austerity-based cuts to social spending erode public health. Their findings reveal that in cut-riddled Greece, suicide rates soared by 20% between 2007 and 2009, and new HIV cases rose by 52% in the first half of 2011. Meanwhile Iceland, despite its sharp recession, has maintained social safety nets and the health of its population. A sobering call for democratic, informed choices in response to recession.



Stung! On Jellyfish Blooms and the Future of the Ocean

Lisa-ann Gershwin UNIVERSITY OF CHICAGO PRESS 456 pp. \$27.50 (2013)

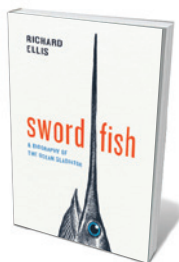
In this bleak take on the future of our seas, Lisa-ann Gershwin chronicles in sometimes exhausting detail how the gelatinous omnivores that are jellyfish are wreaking havoc in waters around the globe. As reams of evidence reveal, human changes to ocean ecosystems are producing perfect conditions for jelly-dominated seas. The switches from chaty to technical language and copious references can jar; but this is a comprehensive summary of the irresistible rise of an arguably unstoppable creature.



The Metamorphoses of Fat: A History of Obesity

Georges Vigarello (translated by C. Jon Delogu) COLUMBIA UNIVERSITY PRESS 296 pp. \$29.50 (2013)

Corpulence and humanity's shifting perceptions of it feature in this curious 'history of the body'. In Jon Delogu's translation, sociologist Georges Vigarello takes us from the twelfth century, when 'fat reduction' was attempted by cutting the limbs with razors to allow the escape of 'wind', to today's confused and often destructive dynamics of thinness and obesity. Vigarello offers up a *grande bouffe* of food for thought, tracing the impact of evolving mores and medicines on society's perception of an often stigmatized condition.



Swordfish: A Biography of the Ocean Gladiator

Richard Ellis UNIVERSITY OF CHICAGO PRESS 272 pp. £18 (2013)

Blue-eyed, bulky, fast and sporting a sword-like upper jaw, the broadbill swordfish (*Xiphias gladius*) emerges in marine natural historian Richard Ellis's portrait as an 'apex' predator with real charisma. At home anywhere from the ocean surface to 600 metres down, its unusual adaptability stems partly from specialized thermogenic tissue in its head that keeps its brain warm and protects its central nervous system. A fascinating dip into the history and biology of a seagoing sabre fighter. **Barbara Kiser**



CARLOS CHEN

Q&A Rahul Mehrotra

Elephant man

Architect Rahul Mehrotra builds with social advocacy in mind. His latest project at Hathi Gaon, a village in Rajasthan, India, provides housing for 100 elephants and their mahouts. A professor at Harvard Graduate School of Design in Cambridge, Massachusetts, he talks about urban evolution and 'impatient capital'.



What was your task at Hathi Gaon?

To design very low-cost housing for the elephants that transport tourists to the Amber Fort in Jaipur, Rajasthan, and their keepers — the mahouts. The

mahouts are traditionally Muslims in a right-wing Hindu state, and the state government had been dragging its heels over rehousing them. The mahout–elephant relationship is complex and close — the mahout often sings the elephant to sleep — so they had to be housed together.

What was your strategy?

Animal-rights groups had put pressure on the government, because the elephants didn't have access to water. Elephants bond with their mahouts through bathing. The elephants are also decorated with toxic paint that needs to be washed off. My firm and I decided that water was key. The land allocated to us was an old sand quarry — essentially a big hole in the ground. We envisaged housing around water pools, filled by the monsoon,

and a slow regeneration of the landscape, over 10 to 15 years. To begin with, the government was not keen to wait that long.

What triggered the go-ahead?

The monsoon. About 65% of the rainwater that filled the pools was absorbed, but that was good because it flooded the local water table and supported our tree-planting project. The clay walls of the pools then compacted naturally, as local craftsmen assured us they would, and now the loss is nearer 25% — enough to keep the pools filled and to support vegetation. Within two years, the landscape was transformed. The government could see what we were trying to do and became enthusiastic about the project.

What are the houses like?

They are small, one-storey structures in which the elephant occupies a portion of the ground floor, and the mahout and his family another. We designed the roofs as flat slabs so the families could build on the upper levels as their incomes grew. The walls are made of local stone. Our most important tools were local wisdom and local materials. The first families arrived in early 2012. Altogether, the village will house between 70 and 100 families.

Can architecture solve social problems?

It can contribute to the solutions. The mahouts had never lived together, so they had no real sense of community. At Hathi Gaon, we were allocated only 45 square metres per family, but we were able to build extra space into each house in the form of inner courtyards. Clusters of houses share bigger courtyards. Communities are now forming.

You describe Mumbai as your lab. Why?

Mumbai embodies many of the problems of expanding cities. City expansion can happen in two ways — involution and evolution. With involution, a space is used in increasingly dense and complex ways. Slums grow in this way, with more and more people and activities packed into the same space. Evolution is expansion through appropriation of surrounding space such as suburbs: the city becomes more diverse. The most successful cities, such as New York, strike a balance between the two. Mumbai has an extreme form of involution: the slums are major centres of production. But it also has evolution potential — a large metropolitan hinterland that it has yet to integrate.

So cities can look similar but differ radically?

Yes. The forms may be similar, but the processes that drive them are unique to each. Although Dubai resembles New York in some ways, it has grown differently. New York was essentially built by its citizens — its form was shaped by human needs. Dubai attracts impatient capital — money from outside interests such as global corporations, translated quickly into new buildings. Providers of curtain glazing tend to have more of an impact on such urban landscapes than citizens do.

NICK HIGGINS

How does that affect those citizens?

Detrimentally. Impatient capital has no long-term commitment to the city. It can distort land values or leave incompatible structures adjacent to one another. An example is gated communities — tower blocks in the inner city or sprawling suburban compounds on the periphery. Impatient capital can be dangerous if it drives planning decisions in a city such as Mumbai, which aspires to be global but isn't yet, because those decisions tend to lose sight of human beings.

Is architecture leading or following the evolution of urban complexity?

This is a live debate in the profession. One school feels that architecture should be avant-garde. Another thinks that it should be responsive to society's problems. But there is a middle ground, where architects can anticipate social problems and reduce or avert them through the imaginative use of space — as we did at Hathi Gaon. ■

INTERVIEW BY LAURA SPINNEY

Correspondence

Social responsibility for new technologies

A 2013 report by the European Environment Agency analyses a range of chemical and technological innovations and the long-term effectiveness of policies designed to minimize risks to health and the environment resulting from their use (see go.nature.com/ajxkkt). As contributors to the report, we call for more proactive forms of governance to incorporate greater social responsibility in new technologies.

The report shows that, rather than stifling innovation, preventive policies can promote investment in safer alternatives and in managing risk. When precautionary regulatory action was taken on the basis of plausible yet incomplete evidence of harm, cases of over-regulation were rare. More often, such action was later shown to be justified.

Twenty case studies of potential hazards — including leaded petrol, nicotinoid pesticides (recently banned in Europe) and mercury pollution — reveal that current risk-assessment standards often prevent or delay the detection of risks. Given the number of emerging technologies that will require safety appraisal, future risk assessments will need to adopt solutions-oriented strategies that increase timeliness and efficacy.

The complex issues around emerging technologies, such as genetically modified crops, call for greater social responsibility by scientists, and for regulators to seek advice from the wider community to address safety concerns and establish principles for sustainability, utility and acceptable risk early on.

Incentives and fiscal advantages, such as public procurement, will encourage industry to adopt responsible measures. These will require social criteria to be captured in processes for public research funding, backed by an updated definition of scientific excellence.

Jacqueline McGlade
*European Environment Agency,
Copenhagen, Denmark.*
jacqueline.mcglade@eea.europa.eu
David Quist *GenØk, Tromsø,
Norway.*
David Gee *London, UK.*

Preserve our digital heritage

Last month saw the sudden end of Yahoo! Messages, a valuable 15-year-old treasure trove of early Internet discussions. Luckily, Archive Team, a group dedicated to saving our digital heritage, was able to preserve the data. But we need to be more vigilant over the erosion of our digital archives (see also *Nature* **493**, 19–21; 2013).

For instance, Yahoo!'s closure of GeoCities in 2009 was described by the Internet Archive as destroying “the most amount of history in the shortest amount of time, certainly on purpose, in known memory”.

Archives of paper documents have been preserved for centuries, and their destruction promotes a strong response (see, for example, go.nature.com/gstbwk). Ditching online archives fails to stimulate similar reactions.

Researchers should publicly question the deletion of digital archives. We can all take part by contributing to the Internet Archive or Archive Team Warrior initiatives, or by learning more through sites such as the digital preservation blog of the US Library of Congress.

Ian Milligan *University of
Waterloo, Ontario, Canada.*
i2milligan@uwaterloo.ca

An innovation prize for clean cookstoves

A radical shift in engine technology in the 1970s (Honda's CVCC) drastically cut motor-vehicle emissions. A comparable game-changer could solve an even bigger pollution problem today.

Household air pollution from the traditional biomass-burning

stoves used in many developing countries is the world's largest environmental-health threat, leading to 4 million premature deaths annually (S. S. Lim *et al. Lancet* **380**, 2224–2260; 2012). We propose that a multimillion-dollar innovation prize should be set up, funded by governments or private philanthropy, to rapidly bring cleaner, more efficient and affordable stoves to poor people.

The competition would attract the world's best combustion scientists and engineers, and would help to take the current efforts of non-governmental organizations, small companies and academics to the next level (see S. Anenberg *et al. Nature* **490**, 343; 2012).

The prize would be awarded for a durable, low-emission biomass-combustion unit — the ‘heart of the hearth’ — rather than for the stove itself. Stove designs could then be adapted around this to meet local requirements.

Ambuj D. Sagar *Indian Institute
of Technology Delhi, India.*

Kirk R. Smith *University of
California, Berkeley, USA.*
krksmith@berkeley.edu

No mass strandings since sonar ban

Prompt political action may have resulted in a remarkable conservation success for whales and dolphins. The Canary Islands used to be a hotspot for mass strandings, but there have been no mass beachings since the Spanish government imposed a moratorium on naval exercises in these waters in 2004.

Naval sonar operations have long been implicated in mass strandings, particularly of beaked whales (R. Filadelfo *et al. Aquat. Mamm.* **35**, 435–444; 2009). In 2002, the stranding of 14 beaked whales in the Canary Islands was linked to the use of mid-frequency naval sonar in the area (P. D. Jepson *et al. Nature* **425**, 575–576; 2003). In 2004, four Cuvier's beaked whales (*Ziphius cavirostris*) were stranded there

after an international naval exercise (A. Fernández *et al. J. Mar. Sci. Res. Dev.* **2**, 2; 2012).

These events prompted the Spanish moratorium and led the European Parliament to issue a non-binding resolution in 2004 to stop the deployment of high-intensity sonar until the completion of a global assessment of its cumulative effects on marine life.

Antonio Fernández, Manuel Arbelo *University of Las Palmas de Gran Canaria, Spain.*
afernandez@dmor.ulpgc.es
Vidal Martín *Society of Cetacean Studies in Canary Islands (SECAC), Lanzarote, Spain.*

Japan's work on ALMA telescope

Your report on the Atacama Large Millimeter/submillimeter Array (ALMA), the world's highest-altitude radio telescope, omits mention of Japan's contribution (*Nature* **495**, 156–159; 2013).

The National Astronomical Observatory of Japan (NAOJ) is one of three executive partners of ALMA and is in charge of implementing East Asia's contribution to the array.

Japan did not simply “join the partnership in 2004” (see D. Normile & D. Clery *Science* **333**, 1820–1823; 2011). As long ago as 1983, Japan put forward plans for the Large Millimeter Array (LMA), almost at the same time as the US National Radio Astronomy Observatory proposed the Millimeter Array. Japan was the first to suggest expansion of the observation wavelength to the submillimetre range in 1987, renaming the LMA as the Large Millimeter and Submillimeter Array (LMSA). It was also the first country to start site surveys for millimetre/submillimetre-wavelength interferometers around the current ALMA site in northern Chile in 1992.

Masahiko Hayashi, Satoru Iguchi *NAOJ, Tokyo, Japan.*
masa.hayashi@z.nao.ac.jp

Robert Edwards

(1925–2013)

Pioneer of *in vitro* fertilization.

Several scientists have made discoveries that have saved millions of lives. Robert Edwards helped to create them.

Edwards, who died on 10 April, was born in 1925 in Batley, UK, a West Yorkshire mill town, and educated in Manchester. He studied agriculture and zoology at the University of Wales, in Bangor, UK, after nearly four years of military service. In 1951, he graduated with only a pass. Despite this inauspicious start, his friend John Slee remembered much later that Edwards had been “ambitious and flexible, and unusually confident in his own judgement”.

Soon after Edwards had enrolled at the University of Edinburgh, UK, to pursue a diploma in animal genetics, his professor, Conrad Waddington, offered him a PhD studentship and later a fellowship. At Edinburgh, in collaboration with fellow graduate student Ruth Fowler (Edwards’ future wife and granddaughter of the eminent physicist Ernest Rutherford) Edwards established how to manipulate ovulation in mice. Six years on, with a bundle of research papers, including several of the many that would be published in *Nature*, Edwards was becoming someone to watch.

After a brief spell at the California Institute of Technology in Pasadena, Edwards was recruited by physiologist Alan Parkes to the Medical Research Council (MRC) National Institute for Medical Research (NIMR) in Mill Hill, London, to help develop vaccines as contraceptives for women. In his spare time, Edwards returned to the study of eggs, primarily wanting to understand the basis of chromosomal abnormalities, as well as what causes eggs to ripen — a necessary first step to achieving *in vitro* fertilization, or IVF.

In the early 1960s, a new NIMR director banned research on human IVF. Following this discouragement, in 1963 Edwards joined Parkes, who had moved to the Physiological Laboratory at the University of Cambridge, UK. Edwards remained at the university for the rest of his life.

At Cambridge, Edwards’ team — which I joined as a PhD student in 1970 — focused on early mammalian development. In a 1965 *Lancet* paper, Edwards had outlined a breathtaking programme for studying the genetics and development of eggs by fertilizing them *in vitro*. To obtain eggs, he needed to collaborate with physicians, the gatekeepers to patients, but his ‘harebrained’ agenda made them wary.



At a now legendary meeting at the Royal Society of Medicine in London, he found the ideal partner. Patrick Steptoe, a senior gynaecologist from Oldham, UK, had completed pioneering work on laparoscopy. Derided by colleagues in gynaecology, this was a surgical technique that Edwards needed to collect eggs from patients. Both men wanted to conquer infertility and both had experienced professional controversy. Their relationship prospered, helped by Jean Purdy, a nurse who was trained by Edwards as a lab technician.

Edwards, Steptoe and a student reported the first convincing evidence of human fertilization *in vitro* in 1969. The clinical implications of the breakthrough predicted by Edwards could no longer be dismissed as pie in the sky. Media attention reached new heights, and a bitter tide of criticism flooded in from scientists, physicians and politicians.

After the MRC declined Edwards’ and Steptoe’s grant application, ostensibly because of concerns about safety, their programme continued to be stretched between Oldham and Cambridge on a shoestring budget. There were still so many unanswered questions. Would fertility drugs be needed? Would the resulting embryos be healthy? Edwards and Steptoe laboured for nearly a decade without a viable pregnancy, but in 1978, Louise Brown, the world’s first ‘test-tube’ baby was born. After her birth, most of the hostility vaporized; it seemed heartless to say that a bonny IVF baby should never have been born. There are now more than five million such babies, many of whom are parents themselves.

In 1980, Edwards and Steptoe founded a private fertility clinic at Bourn Hall, outside Cambridge. Edwards wanted to make IVF acceptable but also, as the father of five daughters, to speak up for people who were infertile. Edwards was their champion in ethically charged battles with scientists, theologians, politicians and even Nobel laureates, whose pantheon he later joined. His hopes were vindicated, but success had come at the price of being accused of everything from killing embryonic ‘babies’ to courting the media.

Our beloved professor could be maddening when he rolled out ideas for experiments like a newspaper press. The late grande dame of embryology, Anne McLaren, a contemporary of Bob’s, once told me: “From scores of ideas, some gems he digs up sparkle so brilliantly they take your breath away.” Bob gave his students freedom to explore, and in an egalitarian environment we flourished.

Still full of energy after retiring in 1989, Bob became a founder–editor of several major journals and continued to oversee *Reproductive BioMedicine Online* until past the age of 80. He lived long enough to enjoy seeing his controversial programme enter mainstream medicine. IVF has led to the banking and donation of eggs and embryos, the ability to genetically diagnose embryos before they are implanted in the uterus, treatment for male infertility, and human embryonic stem-cell technology, which Bob had anticipated in the 1960s. Assisted reproductive technology has transformed the very definition of ‘family’.

Last year, Bob’s papers were on display next to the famous handbag of former British Conservative prime minister, Margaret Thatcher, at his college in Cambridge (Churchill). He and Thatcher were born and died within days of each other and, although political opposites, both stubbornly held visions of a different world, which they strove to make a reality. By 1990, Thatcher’s government had introduced legislation that made IVF widely available. It would be another 20 years and the Nobel Prize in Physiology or Medicine, which was awarded to Bob in 2010, before he was honoured with a knighthood. ■

Roger Gosden is former director of reproductive biology at Weill Cornell Medical College, New York, USA. He was a research student and fellow under Edwards from 1970–76.
e-mail: roger.gosden@cantab.net

CORBIS O’GRADY STUDIO/SPL

FISHERIES

Climate change at the dinner table

An innovative use of catch statistics shows that climate change has already influenced the composition of species in fisheries around the world, and thereby the fish that we eat. [SEE LETTER P.365](#)

MARK R. PAYNE

Fisheries scientists tend to view climate change as a dark cloud on the horizon: potentially problematic in the future, but not of immediate concern. Over the multi-decadal to centennial scale, warming, acidification and deoxygenation of the oceans are expected to have significant impacts on marine ecosystems and fisheries¹. By contrast, other problems, such as the global overfishing crisis², are immediate and pressing and have rightly garnered the field's attention for the past decade. However, on page 365 of this issue, Cheung *et al.*³ present startling evidence that global fisheries catches have already changed in a manner associated with the warming trend — climate change is suddenly an unexpected guest at dinner.

Changes in the spatial distribution of species are one of the major predicted impacts of climate change on marine ecosystems⁴. Marine species tend to occupy the full range of temperatures that they can physiologically tolerate and therefore, on a broad scale, their distributions closely track temperature boundaries in the ocean⁵. Distribution changes generally occur over multiple generations through the waxing and waning of local populations: warming can alter the growth, survival or reproduction rates of individuals, favouring warm-adapted species and disadvantaging cold-adapted ones^{5,6}. Thus, all other factors being equal, warming is expected to change the species composition of fish communities in a given region, replacing colder-water species with warmer-water ones.

A subtle but important difference exists in tropical waters. Because these regions are generally the hottest parts of the ocean, there are simply no species from hotter climes. Initial warming is expected, as elsewhere, to shift community composition in favour of warm-adapted species. However, once the cooler-water species have declined, the community composition is predicted to stabilize, owing to the lack of hotter-water species that could expand into these regions. The composition will thereafter remain stable, even as these regions become too hot for the remaining warm-adapted species, and their numbers also decline⁵.



ROS DRINKWATER/ALAMY

Figure 1 | Mullet on the move. Cheung and colleagues' study³ shows that the catch composition of fisheries has been changing worldwide since the 1970s, as average water temperatures have increased. Red mullet (*Mullus surmuletus*), for example, shown here on sale in a Portuguese market, is usually associated with the Mediterranean, but has in recent decades expanded into the North Sea and Norwegian waters¹¹.

Demonstrating these hypotheses on a large scale, however, has proved problematic. Measurements of fish distributions are typically based on resource-intensive scientific studies, such as sonar estimates and trawl surveys, and are usually limited to the developed world. Catch statistics, by contrast, are ubiquitous and routinely gathered in developed and developing economies. But their use is highly contested (see, for example, a recent Comment in *Nature*⁷) and they would not normally be considered suitable for this type of study.

Cheung and colleagues took an innovative approach to address this problem by analysing the mean temperature of the catch (MTC). This ambiguously named metric does not refer to the temperature of the fish as they leave the water, as the name implies, but instead characterizes the composition of the catch in terms of the 'thermal preferences' of species⁸. For each species in their database, the authors derived

a characteristic temperature preference from species-distribution modelling. The MTC was then calculated as the catch-weighted mean of these thermal preferences.

The authors applied this metric to the annual catch of 990 species across 52 large marine ecosystems from 1970 to 2006. After accounting for potential confounding factors, they discovered a significant positive correlation between the regional trend in sea surface temperature and that in the MTC. In tropical ecosystems, sea surface temperatures have warmed systematically, but the MTC stabilized after an initial increase in the 1970s. The patterns are seen globally, are independent of other forms of oceanographic variability and fishing intensity, and are consistent with the proposed hypotheses. Importantly, the method used to define the thermal preferences used in calculating the MTC does not seem to influence the result.

These findings therefore make a strong case

that warming sea surface temperatures are changing catch composition around the world (Fig. 1). Furthermore, they suggest that catch composition in tropical regions has stabilized, implying that these regions have already become too hot for all but the warmest-water species.

The key to this approach, and what separates it from other studies, is how the catch data are used. Global catch data are most commonly used to assess the state of the world's fish populations (for example, see refs 2 and 9). Cheung and colleagues, however, use catch data to infer the species composition of fisheries and how they have changed over time; the data are therefore directly relevant to the question asked, rather than an indirect measure. As such, many of the criticisms levelled at other catch-based studies do not apply in this situation.

There is a strong temptation to over-interpret this result and conclude that the underlying marine populations have also changed in response to warming — in addition to the composition of fisheries catches. The authors largely resist this temptation, although they present some evidence that it is a valid conclusion in at least one region, the North Sea. However, fisheries are influenced by many processes beyond the biological realm, including consumer preferences, technological advances, international politics and fuel costs. Catch therefore does not always reflect underlying abundance⁷. The impact of these factors on the MTC metric needs to be understood before generalizations about the underlying fish communities can be made.

Nevertheless, the changes in the make-up of the catch alone are startling and cause for concern, particularly in tropical countries. In these regions, the fact that the catch composition seems to have reached the terminal hot-water state means that further warming may reduce fishery yields, if it has not already done so. The countries that border tropical waters are also those that are the most dependent on fisheries as sources of employment, foreign revenue and food, and are least able to adapt to such changes¹⁰. Shifts in the composition of their fisheries, and potential reductions in fishery productivity, are therefore likely to affect these states' food security and development.

The changes reported by Cheung and colleagues are global in nature. Consumers and fishing communities will need to adapt to new species appearing in fishing nets and the dwindling presence of traditional species. Resources will be required to aid this adaptation, particularly in tropical countries. Climate change has made it to the fishmonger and onto our dining tables. The question now is, how should we respond? ■

Mark R. Payne is at the Centre for Ocean Life, National Institute for Aquatic Resources

(DTU Aqua), Technical University of Denmark, Charlottenlund 2920, Denmark.
e-mail: mpa@aqua.dtu.dk

1. Gruber, N. *Phil. Trans. R. Soc. Lond. A* **369**, 1980–1996 (2011).
2. Worm, B. *et al. Science* **325**, 578–585 (2009).
3. Cheung, W. W. L., Watson, R. & Pauly, D. *Nature* **497**, 365–368 (2013).
4. IPCC. *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the*

Intergovernmental Panel on Climate Change 976 (Cambridge Univ. Press, 2007).

5. Sunday, J. M., Bates, A. E. & Dulvy, N. K. *Nature Clim. Change* **2**, 686–690 (2012).
6. Drinkwater, K. F. *et al. J. Mar. Syst.* **79**, 374–388 (2010).
7. Pauly, D., Hilborn, R. & Branch, T. A. *Nature* **494**, 303–306 (2013).
8. Collie, J. S., Wood, A. D. & Jeffries, H. P. *Can. J. Fish. Aquat. Sci.* **65**, 1352–1365 (2008).
9. Costello, C. *et al. Science* **338**, 517–520 (2012).
10. Allison, E. H. *et al. Fish Fish.* **10**, 173–196 (2009).
11. Beare, D., Burns, F., Jones, E., Peach, K. & Reid, D. *J. Sea Res.* **53**, 205–212 (2005).

MICROBIOLOGY

Bacterial communities as capitalist economies

Tracking the behaviour of bacteria as they group together on a surface reveals a 'rich-get-richer' mechanism in which polysaccharide deposition and cellular location amplify in a positive feedback loop. SEE LETTER P.388

UTE RÖMLING

Multicellular communities of single-celled organisms attached to a surface are the predominant form of life on Earth. Development of these biofilms involves distinct stages of self-organization, starting with a single cell that senses and approaches a surface. Numerous factors have been identified that affect the formation, structure, metabolism and regulation of biofilms at the population level, but their formation has rarely been quantitatively investigated at the single-cell level. On page 388 of this issue, Zhao *et al.*¹ track the fate of individual bacterial cells from reversible surface attachment to microcolony formation. They show that the early stages of cell-to-surface contact are highly dynamic, with the bacteria exploring the surface and priming their environment for subsequent biofilm development*.

During the early stages of biofilm formation², motile bacteria approach a surface using a swimming action mediated by flagella — helical rotatory filaments that are more than twice the cell's length. On sensing a surface, a tighter, although reversible, contact is established through retractable appendages such as pili or fimbriae, which support surface crawling and contribute to the shaping of the three-dimensional architecture of biofilms³. Surface contacts mediated through these cellular extensions subsequently trigger the bacteria to secrete adhesive extracellular-matrix components^{4,5}, which mediate irreversible binding, a prerequisite for the development of microcolonies on biotic and abiotic surfaces. These microcolonies (clusters of up to 50 cells) are

the foundation of the mature three-dimensional biofilm, which has distinct physiological properties, including antibiotic resistance and immune-system resilience.

But exactly how surface contact is translated into microcolony formation is not fully understood. To address this question, Zhao *et al.* explored the fate of individual cells of *Pseudomonas aeruginosa*, an environmental bacterium and formidable pathogen. They tracked hundreds of thousands of cells in a defined area in time and space using a massively parallel cell-tracking algorithm⁶. The authors found that the cells used type IV pili to explore the surface immediately after initial contact. Surprisingly, the bacteria did this in a non-random manner: a significant area of the surface was never visited, many places were visited once, and a few locations were visited more than 100 times.

To investigate this unexpected result, Zhao and colleagues focused on the role of Psl, the adhesive polysaccharide secreted by *P. aeruginosa*⁵. Their experiments revealed that the bacteria leave a trail of Psl as they move along the surface, but not of extracellular DNA, another component required for the initial establishment of biofilms. The researchers also found that Psl was unequally distributed on the surface — the areas with highest Psl deposition corresponded to those that had been visited the longest by bacteria, through either multiple visits or longer visiting time, and these sites constituted the founding sites of microcolonies. The authors propose that the mediators of the unequal surface exploration by the bacteria are the type IV pili, and the fact that these are attracted to Psl-rich regions generates a positive feedback loop.

*This article and the paper under discussion¹ were published online on 8 May 2013.

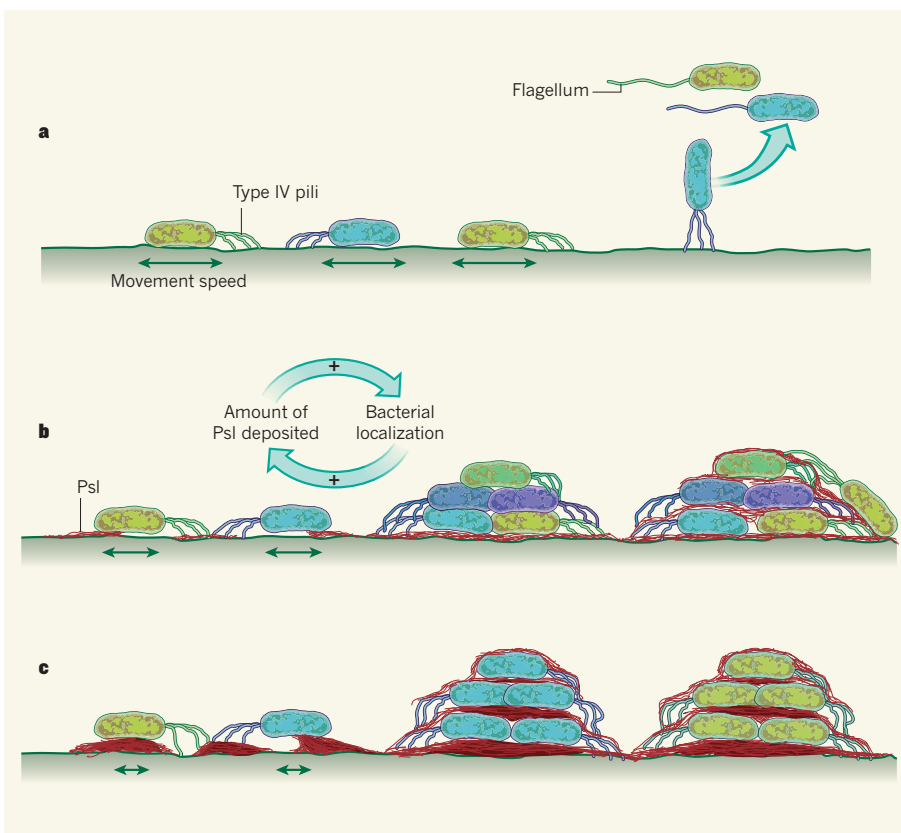


Figure 1 | A positive feed-forward circuit steers biofilm formation. *Pseudomonas aeruginosa* bacteria use retractable structures called type IV pili to crawl along a surface after initial contact. Zhao and colleagues¹ show that the cells' exploration of the surface and subsequent development of a multicellular biofilm depends on their secretion of the adhesion molecule Psl, the presence of which promotes bacterial residence at a particular site. **a**, Mutant cells that produce no Psl explore a large surface area but are unable to irreversibly attach to it, such that no microcolonies form. **b**, Wild-type *P. aeruginosa* leave trails of Psl as they explore the surface. In a positive feedback loop, Psl deposition extends the cells' sojourn time, such that they self-organize and form microcolonies at the sites of greatest Psl levels. The daughters of some cell divisions will remain, but because cells can still enter or leave the microcolony, it will contain several cell lineages. **c**, Bacteria that produce high amounts of Psl partake of less surface exploration, and a high proportion of daughter cells are retained at the location of cell division, leading to exponential growth of microcolonies founded by only one lineage.

To test this idea, Zhao *et al.* generated a mutant *P. aeruginosa* strain that lacked the gene encoding Psl. Although this mutant was unable to initiate irreversible attachment and microcolony formation, the cells still moved along the surface, albeit with significantly shorter contact times (Fig. 1a). Taking the total number of cell-to-surface visits into account revealed that the mutant bacteria explored a larger surface area in a more random way than the wild-type cells. The authors then investigated a *P. aeruginosa* strain that produced higher levels of Psl than the wild-type bacterium: these cells spent yet more time in highly visited areas and even less time elsewhere.

Thus, it seems that the bacteria are guided by a synergistic 'rich get richer' mechanism, in which cells go where other cells go most often. The authors show that the bacterial surface-exploration behaviour could be described by a power law called Zipf's law — a probability distribution in which the frequency of an event decreases with a defined interval. In this case,

the event is the visit frequency, and sites visited only once are most numerous, sites visited twice are ranked next, and so on. Many self-organized systems, including wealth distribution in capitalist economies, follow Zipf's law. From the authors' results, it also follows that the frequency distribution of bacterial visits varies as a power of Psl secretion. Thus, Zhao and colleagues suggest that, for the bacteria, this system results in some 'elite' bacteria being located at sites that are extremely rich in communally produced Psl, and that this social structure is required for microcolony formation.

The observed Psl-dependent movement also had a profound effect on microcolony composition. Microcolonies formed by wild-type cells included several lineages because cells left and joined the area (Fig. 1b). By contrast, in Psl-overproducing cells, microcolonies formed earlier and the bacterial cells were of one lineage, because mother and daughter cells tended to remain in the area (Fig. 1c).

This study and previous work from the

same research group⁶ provide the most detailed picture yet of early biofilm formation in *P. aeruginosa*. The present results also define the phase between reversible and irreversible attachment — during which both motility and adhesion factors are expressed, enabling the bacteria to scan the surface of a potential new home and simultaneously prepare the home for settlement — as a distinct developmental stage in biofilm formation. It will be intriguing to see whether other biofilm-forming bacteria, such as *Escherichia coli*⁷ or *Salmonella typhimurium*⁸, connect surface movement and polysaccharide production in a similar way. It will also be interesting to assess how opposite processes — motility and attachment — are simultaneously regulated in a single cell. One obvious candidate to govern both processes is the ubiquitous bacterial signalling molecule cyclic dimeric GMP (c-di-GMP), which regulates the transition between motility and sessility⁹. Psl stimulates the production of c-di-GMP¹⁰ such that the priming of surfaces with Psl gradually directs cells towards sessility, but also exposes individual cells to different microenvironments. This can lead to variability in the population, including bistability.

Surface scanning leading to social self-organization is not unique to *P. aeruginosa* and the initiation of biofilm formation. The gastrointestinal pathogen *S. typhimurium*, for example, uses flagella-mediated near-surface swimming for exploration of the surface of epithelial cells in the gut¹¹. Here, the exploratory movement facilitates target-site selection and cooperative invasion by the bacteria as they infect these cells. Thus, it seems that the onset of biofilm formation and the infection process are governed by similar basic principles, suggesting that this behaviour may be a target for developing new ways to inhibit surface colonization by bacteria or to prevent infection. ■

Ute Römling is in the Department of Microbiology, Cell and Tumor Biology, Karolinska Institutet, 171 77 Stockholm, Sweden.
e-mail: ute.romling@ki.se

1. Zhao, K. *et al.* *Nature* **497**, 388–391 (2013).
2. O'Toole, G., Kaplan, H. B. & Kolter, R. *Annu. Rev. Microbiol.* **54**, 49–79 (2000).
3. Barken, K. B. *et al.* *Environ. Microbiol.* **10**, 2331–2343 (2008).
4. Li, G. *et al.* *Mol. Microbiol.* **83**, 41–51 (2012).
5. Mann, E. E. & Wozniak, D. J. *FEMS Microbiol. Rev.* **36**, 893–916 (2012).
6. Gibiansky, M. L. *et al.* *Science* **330**, 197 (2010).
7. Thomas, W. E., Nilsson, L. M., Forero, M., Sokurenko, E. V. & Vogel, V. *Mol. Microbiol.* **53**, 1545–1557 (2004).
8. Grantcharova, N., Peters, V., Monteiro, C., Zakikhany, K. & Römling, U. *J. Bacteriol.* **192**, 456–466 (2010).
9. Pultz, I. S. *et al.* *Mol. Microbiol.* **86**, 1424–1440 (2012).
10. Irie, Y. *et al.* *Proc. Natl Acad. Sci. USA* **109**, 20632–20636 (2012).
11. Misselwitz, B. *et al.* *PLoS Pathog.* **8**, e1002810 (2012).

PLANETARY SCIENCE

Plumbing the depths of Uranus and Neptune

An analysis of data collected by the Voyager 2 spacecraft and by ground-based telescopes limits the depths to which winds penetrate into Uranus and Neptune, informing the debate about these planets' internal structures. [SEE LETTER P.344](#)

PETER READ

In an age in which space missions penetrate the far reaches of the Solar System, it seems incredible that planetary scientists still argue about the depth of atmospheric circulations inside the ice giants, Uranus and Neptune, and the gas giants, Jupiter and Saturn. But, because of the complexities of the problem, many uncertainties remain. The difficulties are compounded by the inaccessibility of the planets' interiors, which are hidden beneath dense layers of clouds, making it difficult to observe their structure and dynamics. It is therefore a great achievement for Kaspi *et al.*¹ (page 344 of this issue) to have obtained, from careful and insightful analyses of the planets' gravitational fields, surprisingly tight constraints on how deeply the winds on Uranus and Neptune penetrate.

How do we know anything about the winds on these remote planets? The main sources of information date back to the close encounters of the Voyager 2 spacecraft with Uranus (in 1986) and Neptune (in 1989) during its grand tour of the outer Solar System. Images obtained during those encounters revealed methane and ammonia clouds whose motion could be tracked to determine wind strength at levels at which the atmospheric pressure was around 100–300 kilopascals. This showed that, despite being among the farthest flung and coldest planets in the Solar System, Uranus and Neptune actually have some of the strongest winds², with east–west (zonal) velocities of up to 450 metres per second, compared with just 30–100 m s⁻¹ on Earth. The source of energy for these immensely strong winds remains mysterious — it is unclear whether they are driven mainly by weak differential heating from the Sun or, at least in the case of Neptune, by heat upwelling from the deep interior.

Uranus and Neptune are composed primarily of hydrogen, helium and so-called icy materials (water, ammonia and hydrocarbons, albeit at temperatures of several thousand kelvin), so the interior of both planets is almost certainly predominantly fluid^{3,4}. This raises the question of whether the winds observed at the 100-kPa level are confined to the outer atmospheric layers, or represent a deep-seated pattern of internal thermal convection. But

how can researchers tell the difference? Direct measurements are limited to the outermost few tens of kilometres of the atmosphere, where cloud motions can be tracked. Indirect methods, such as those based on inferences and insight from numerical simulations, must therefore be adopted. However, these simulations have severe limitations because it is impossible to model conditions that even remotely approach those that prevail⁵ in the deep interiors of Uranus and Neptune.

On the basis of principles established by the fluid dynamicists Geoffrey Ingram Taylor and Joseph Proudman^{6,7}, Fritz Busse noted⁸ that fairly weak differential motions in a rapidly rotating fluid tend to be coherent (almost invariant) along the direction parallel to the rotation axis, as long as surfaces of constant density are approximately horizontal. Inside a spherical planet, motions would therefore tend to be almost constant on cylindrical surfaces coaxial with the planetary rotation axis, if the fluid density were the same everywhere (Fig. 1a). In the presence of purely radial variations in density ρ , however, the

product of ρ and the velocity U in the direction perpendicular to the rotation axis must be coherent along cylinders. So U would become much smaller inside the planet as the density increases, reducing by a factor of about 10,000 between the 100-kPa level and the deep interior of Uranus or Neptune (Fig. 1b).

If latitudinal variations in density are included, the strict axial coherence of the flow can be broken by the effects of density variations along horizontal surfaces (known as a baroclinic flow). Real convective flows are likely to be at least partly baroclinic. Kaspi and colleagues have extended an earlier method⁹ by taking into account the density variations associated with baroclinic effects to constrain the depth to which the surface winds on Uranus and Neptune penetrate.

Variations in density that balance the variations in ρU along cylindrical surfaces will also affect the structure of a planet's gravitational field, which can be measured from small accelerations and decelerations in the trajectory of a spacecraft as it flies close to the planet. The zonal winds of Uranus and Neptune vary only gradually with latitude², therefore the effects of dynamic density variations should manifest themselves in the low-order gravitational harmonics that represent the largest-scale variations in the planets' gravitational fields. Two of these harmonics (designated J_2 and J_4) have been measured, for example during the encounters of Voyager 2 with Uranus and Neptune. Kaspi *et al.* have determined that only a velocity pattern that is tightly confined to the outermost 1,000 km of each planet's radius can be consistent with the measured J_2 and J_4 harmonics. This result probably favours a shallow meteorology for both planets

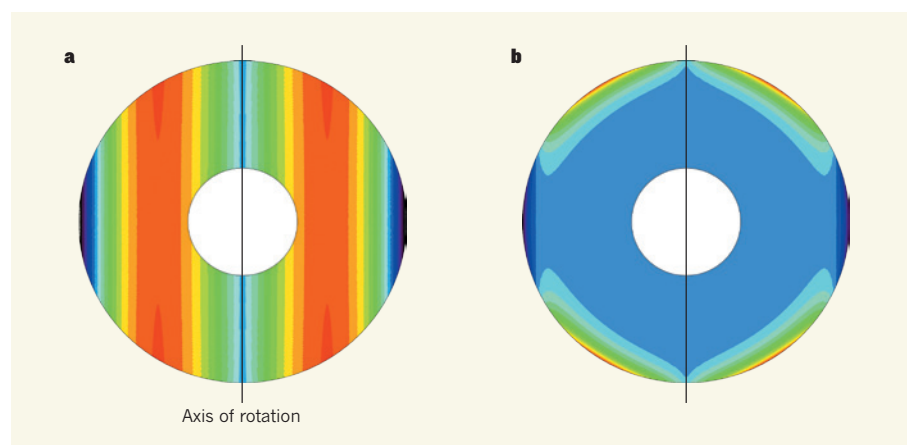


Figure 1 | Velocity fields in fluid planets. In these cross-sections through a fluid planet that rotates about a vertical axis, the colour contours represent the magnitude of a hypothetical zonal-wind field, describing fluid motion for which $2\Omega \cdot \nabla(\rho U) = 0$, where Ω is angular velocity of planetary rotation, ρ is fluid density, assumed to vary exponentially with radius, and U is fluid velocity in the east–west direction. Dark blue, westward flow; red–orange, eastward flow; yellow, weak-to-moderate eastward flow; pale blue and green, flow close to zero. The cross-sections represent examples of the kinds of interior flow pattern considered by Kaspi *et al.*¹ for Uranus and Neptune. In each case, the latitudinal surface-velocity pattern is consistent with that of those planets. **a**, The density of this planet is essentially uniform. The velocity field therefore hardly varies along coaxial cylinders around the axis of rotation. **b**, The variation of density with depth for this planet is roughly the same as that of Uranus, increasing inwards by a factor of around 10,000 from the visible cloud tops. The velocity field strongly intensifies at the surface.

(although 1,000 km seems deep to us on Earth, because Earth's atmosphere spans a depth of just 100 km).

So what does this suggest for the larger cousins of these planets, Jupiter and Saturn? They also have strong zonal winds at their cloud tops, but these winds vary much more strongly with latitude than those of Uranus and Neptune. This means that any signature in the structure of their gravitational fields will manifest itself only in the high-order, rapidly varying gravitational moments (J_{10} and higher)¹⁰, which could not be measured during the Voyager fly-bys or in the ongoing Cassini mission.

In 2011, NASA launched its Juno mission, the main goal of which is to accurately measure the high-order gravitational moments of Jupiter from the vantage point of a low-altitude polar orbit. Kaspi and colleagues' approach will enable constraints to be placed on the depth of Jupiter's deep baroclinic flow. This might settle the long-running debate over whether Jupiter's winds are deep-seated or superficial. And

when the Cassini orbiter reaches the end of its mission in 2017, plans are afoot for the spacecraft to enter a low polar orbit around Saturn (before crashing into it) to make the same measurements of the planet's gravitational field as Juno will for Jupiter. By the end of this decade, we may therefore have obtained measurements that constrain the wind-penetration depth for all four gas- and ice-giant planets.

If I were a betting man, my money would be on a shallow penetration depth for the gas giants, given the precedents of their ice-giant cousins. There are several other reasons to expect gas-giant meteorology to be confined to shallow depths¹¹, one of which is that the fluid interiors of the gas giants become electrically conducting at depths of several thousand kilometres. This allows hydromagnetic forces to disrupt the 'Taylor–Proudman condition', breaking the associated axial coherence of the winds. However, a probe from the Galileo spacecraft found that zonal winds increase with depth near Jupiter's equator¹², so the evidence is

not unequivocal. I await the results from Juno and Cassini with considerable interest. ■

Peter Read is in the Department of Physics, Clarendon Laboratory, University of Oxford, Oxford OX1 3PU, UK.

e-mail: p.read1@physics.ox.ac.uk

1. Kaspi, Y., Showman, A. P., Hubbard, W. B., Aharonson, O. & Helled, R. *Nature* **497**, 344–347 (2013).
2. Ingersoll, A. P. *Science* **248**, 308–315 (1990).
3. Hubbard, W. B. & Marley, M. S. *Icarus* **78**, 102–118 (1989).
4. Podolak, M., Weizman, A. & Marley, M. *Planet. Space Sci.* **43**, 1517–1522 (1995).
5. Showman, A. P., Kaspi, Y. & Flierl, G. R. *Icarus* **211**, 1258–1273 (2011).
6. Taylor, G. I. *Proc. R. Soc. Lond. A* **93**, 92–113 (1917).
7. Proudman, J. *Proc. R. Soc. Lond. A* **92**, 408–424 (1916).
8. Busse, F. H. *Icarus* **29**, 255–260 (1976).
9. Hubbard, W. B. *et al.* *Science* **253**, 648–651 (1991).
10. Hubbard, W. B. *Icarus* **137**, 357–359 (1999).
11. Liu, J., Goldreich, P. M. & Stevenson, D. J. *Icarus* **196**, 653–664 (2008).
12. Atkinson, D. H., Ingersoll, A. P. & Seiff, A. *Nature* **338**, 649–650 (1997).

water, or by the electrolysis of water³. Either way requires two steps: hydrogen generation and the reduction of the iron oxides. It therefore seems logical to bypass the hydrogen-generation step and use electricity to reduce iron ore directly (Fig. 1). What is more, it would be preferable to use conditions in which both the metal product and the electrolyte (the material in which the iron oxide is dissolved) are liquids, because liquids are generally easier to handle than solids. Another important reason for making a liquid product is that if iron is deposited as a solid from molten electrolyte, it forms as a fine powder that can be easily oxidized, reversing the desired reaction. The overall challenge, therefore, is to devise an electrolytic process that produces carbon-free liquid iron and oxygen.

Iron oxide readily dissolves in metal-oxide melts to form a predominantly ionically conducting mixture in which the iron is present as

METALLURGY

Iron production electrified

Scientists have long dreamt of converting molten iron oxide to iron and oxygen using electricity. An anode material that withstands the high temperatures and corrosive chemicals involved brings the dream closer to reality. SEE LETTER P.353

DEREK FRAY

In 2011, about one billion tonnes of iron was produced worldwide¹. Unfortunately, the concomitant generation of carbon dioxide contributed to around 5% of that year's global increase in atmospheric CO₂ (ref. 1). This was not because the amount of CO₂ produced per tonne of iron is high — it is in fact less than that of most metals — but because of the immense quantity of iron produced. In this issue, Allanore *et al.*² (page 353) report a discovery that might lead to a more environmentally friendly method for making iron: the reduction of iron ore using an electric current, a process known as electrolysis*.

Iron ore is commonly chemically reduced by carbon in a blast furnace at 1,600 °C to give liquid iron that is saturated with carbon, and a mixture of CO₂ and carbon monoxide³. The carbon monoxide is usually burnt to generate heat and more CO₂. However, solidified iron from blast furnaces has few applications because of its inherent brittleness. Useful steel products can be made only by removing most

of the carbon as yet more CO₂ and carbon monoxide, and by removing other impurities that tend to be introduced with the carbon.

An alternative approach is to reduce iron oxides directly with hydrogen. But most hydrogen is obtained by the reaction of methane with



Figure 1 | Metal extraction by electrolysis. The interior of an aluminium smelter demonstrates the harsh conditions that must be withstood by the components of electrolytic cells. Allanore *et al.*² report materials that could be used as anodes in the electrolytic production of iron.

*This article and the paper under discussion² were published online on 8 May 2013.

iron(II) or iron(III) ions. Deposition of liquid metal from the melt is unlikely to be a problem, so the main challenge in developing an electrolytic method for iron production is to find a suitable anode, which must be highly conducting and not be attacked by the melt or by liberated oxygen at 1,600 °C. There are three kinds of material that might be suitable: metals, which are usually highly conducting, but may oxidize; conducting ceramics, which might dissolve in the melt and are not malleable; and cermets, which are mixtures of metals and ceramics⁴.

Allanore *et al.* concentrated on metallic alloys that form oxide films on their surfaces during electrolysis — films that do not readily dissolve in the liquid electrolyte. The same group has also reported⁵ that iridium is a suitable anode material because its oxide cannot form at above 1,200 °C; the metal therefore remains as a metal during electrolysis, rather than corroding away because of oxidation. However, iridium is much too expensive and scarce to use for industrial iron production.

By examining cheaper chromium–iron alloys, the authors have found these to be resistant to oxidation under the conditions that are necessary for iron oxide reduction in molten electrolyte. Another advantage of these alloys is that any anodic dissolution will not introduce undesirable contaminants into the iron. In fact, traces of chromium could be beneficial, because this metal is commonly added to iron to decrease its oxidation rate.

When an alloy oxidizes, the main components of the film that forms at the surface are usually the oxides of the elements in the alloy. Surprisingly, the oxide layer that formed on the anode during electrolysis in Allanore and colleagues' study consisted of a solid solution of chromium oxide and aluminium oxide. The latter came from the electrolyte, which was a mixture of iron(III) oxide, calcium oxide, aluminium oxide and magnesium oxide. But curiously, when the researchers immersed a chromium–iron anode in the electrolyte without an applied voltage, they observed the formation of an additional mixed layer of calcium oxide and aluminium oxide on top of the chromium oxide–aluminium oxide layer. This observation runs counter to conventional oxidation theory. It may be that, when a voltage is applied, positively charged calcium ions are repelled by the positive potential at the anode, preventing calcium from forming part of the coating.

Allanore and co-workers' results will stimulate further development of affordable alloys for iron oxide electrolysis, and will prompt researchers to consider how to design a large pilot reactor. However, considerable technical development will be required for the authors' discovery to be used commercially. For example, electrolytic reactions invariably take place in two dimensions on electrodes, whereas the chemical reactions in conventional

iron-making take place in three dimensions — so it could be difficult to match the space and time yields of conventional smelters in an electrolytic reactor. Electrolytic cells will need to be engineered in which anode-to-cathode distances are short enough to avoid voltage losses in the electrolyte, but not so short that the products (iron and oxygen) formed at each electrode come into contact and react to reform iron oxides. Electrical current densities in the system must also be kept high to maintain a high production rate.

Although there is much to do, a process that is less polluting than existing technology could be developed. This approach could also be applied to the extraction of other metals from their oxides. And there is another exciting potential application: the production of oxygen from metal oxides could be invaluable for space exploration. If the process were

carried out on the Moon, for example, the gas could be used in the fuel–oxygen mixture that is needed for rockets and to sustain life^{6,7}, making human colonization of the Solar System more feasible. ■

Derek Fray is in the Department of Materials Science and Metallurgy, University of Cambridge, Cambridge CB2 3QZ, UK.
e-mail: djf25@cam.ac.uk

1. www.worldsteel.org
2. Allanore, A., Yin, L. & Sadoway, D. R. *Nature* **497**, 353–356 (2013).
3. Habashi, F. *Handbook of Extractive Metallurgy* (Wiley-VCH, 1997).
4. Sadoway, D. R. *J. Metals* **53**, 34–35 (2001).
5. Kim, H., Paramore, J., Allanore, A. & Sadoway, D. R. *J. Electrochem. Soc.* **158**, 101–105 (2011).
6. Sanderson, K. *Nature* <http://dx.doi.org/10.1038/news.2009.803> (2009).
7. Schwandt, C., Hamilton, J. A., Fray, D. J. & Crawford, I. A. *Planet. Space Sci.* **74**, 49–56 (2011).

EVOLUTION

Stuck between the teeth

A computer model of tooth evolution designed to assess the impact of developmental dynamics on natural selection reveals that complexity reduces the likelihood of maximum fitness being attained. [SEE LETTER P.361](#)

P. DAVID POLLY

The evolutionary biologist Leigh Van Valen famously defined evolution as the control of development by ecology¹. By this, he meant that evolution is governed by the fitness of the phenotype of an organism (its physical form); that fitness depends on the ecological context in which the phenotype exists; and that an individual's phenotype is not a static structure built directly from genes, but a dynamic structure that unfolds sequentially from fertilization to adulthood. The developmental sequence determines and constrains which phenotypes can emerge, such that evolutionary change depends on modifications to development. On page 361 of this issue, Salazar-Ciudad and Marín-Riera² present a groundbreakingly realistic computational model of this process. They use their model to demonstrate that developmental complexity often prevents natural selection from reaching optimal fitness when fitness is directly linked to attaining a particular phenotype, but that these 'adaptive peaks' can be reached when fitness is instead linked to functional properties of the phenotype³.

The authors' model revolves around a simulation of the evolution of mammalian tooth development, using well-understood details of molecular developmental biology

and functional ecology^{3–5}. Mammalian teeth develop from two main tissue types, the mesenchyme and the epithelium. The growth of these tissues helps to shape the tooth germ (the aggregation of cells that eventually form the tooth) as it develops through bud, cap and bell stages (Fig. 1a). A suite of activator and inhibitor molecules controls the rates of cell proliferation, differentiation and death in these tissues⁶. Three-dimensional changes in the developing tissues modulate the distances over which these signals interact, causing local changes in their concentration that result in the formation or loss of signalling centres called enamel knots. The shape into which the dynamically controlled tooth germ folds determines the topography of the crown of the mature tooth and thus its functional properties.

A huge variety of teeth can be produced by this developmental system, ranging from a dog's high-cusped molars to the elaborately ridged molars of some rodents, such as the coypu (Fig. 1b). Tall cusps in mammals are usually associated with the puncturing, shearing and tearing functions of carnivores and insectivores, whereas low cusps, crests and basined surfaces are associated with the grinding and chewing functions of omnivores and herbivores⁵.

Salazar-Ciudad and Marín-Riera simulated this developmental system as two sheets of cells, the epithelium and mesenchyme,

^{*}This article and the paper under discussion² were published online on 1 May 2013.

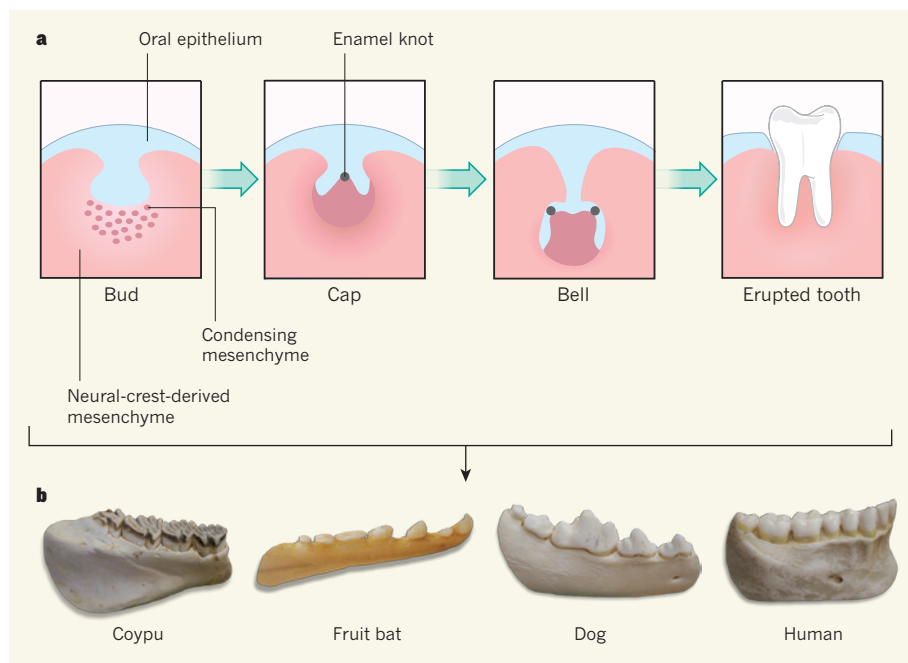


Figure 1 | Mammalian tooth development. **a**, Mammalian teeth develop in the jaws before erupting into the mouth. During this process, the shape of their crown is set by interactions between epithelium and mesenchyme tissues. All teeth start as a bud of epithelium and mesenchyme (bud stage), which grows into a cone-like shape (cap stage). Continued growth and folding of the tissues, through the bell stage to an erupted tooth, is dynamically controlled by molecular signalling centres called enamel knots, which are located where the cusp tips will form. **b**, This developmental system is flexible enough that evolutionary modifications to the expression of the molecular signals can produce a remarkable diversity of tooth shapes. Salazar-Ciudad and Marín-Riera² use this system to model the hypothesis of evolution as the control of development by ecology.

with parameters that represented activator and inhibitor molecules produced by these tissues. Changes in the values of these parameters simulated up- or downregulation of the molecular signals, which, in turn, altered the pattern of growth and folding in the simulated tissue layers, just like in real tooth germs. To simulate evolution of tooth structure using this model, the authors created artificial populations of developing teeth that they subjected to mutation and selection. Each population started with a randomly chosen phenotype. Mutations were applied by stochastically altering the molecular signalling parameters, which resulted in new phenotypes. Fitness was assigned to each individual on the basis of either its phenotypic or its functional similarity to an arbitrarily chosen target phenotype. The latter can be thought of as the phenotype that conveys the greatest fitness in the ecological context of the simulation — a peak on the adaptive landscape, as described by the geneticist Sewall Wright⁷. The fittest individuals were then selected as the parents of the next generation and the simulation was repeated.

This process was expected to cause the population's mean phenotype to evolve towards the model's adaptive peak. And so it did, although only to a point. The authors found that when the criterion for fitness was overall similarity to the optimal phenotype, fewer than 40% of the simulations reached

their adaptive peak. They suggest that failure to reach the adaptive peak is due to the complexity of developmental interactions; the sequence of parameter changes that is needed to generate the optimum phenotype includes changes that temporarily reduce fitness, but selection prevents these changes from occurring, and thus prevents the phenotype from evolving to the adaptive peak. It has long been known that natural selection might leave populations stranded on suboptimal peaks⁸. Indeed, Wright himself introduced the shifting balance theory to explain how an evolving species might jump from a suboptimal peak onto a higher one⁷. Salazar-Ciudad and Marín-Riera have shown that not only are suboptimal dead ends an evolutionary possibility, but they are also exceedingly likely to occur in real, developmentally complex structures when fitness is determined by the exact form of the phenotype.

However, the authors also found that when fitness was determined by functional properties instead of the phenotype itself, the adaptive peak was usually reached. This is because many different phenotypes can have the same functional properties⁹ — a herbivorous mammal, for example, simply needs grinding and chewing surfaces on its teeth, regardless of how the surfaces are constructed. Thus, there are many more paths to a functional adaptive peak than to a phenotypic one, especially for a

phenotype that has a complex developmental system, such as a tooth.

By developing a model of Van Valen's hypothesis of evolution as the control of development by ecology, Salazar-Ciudad and Marín-Riera have demonstrated that one fundamental assumption that underpinned his concept was wrong. Van Valen argued that the effect of genes on evolution is negligible because developmental interactions — through mutation, recombination and epigenetic alteration — are "sufficiently flexible" for any phenotype to be selected¹. By contrast, Salazar-Ciudad and Marín-Riera's results demonstrate that developmental interactions in fact prevent selection from finding the optimal phenotype, even with sufficient mutation in the genetic parameters.

Salazar-Ciudad and Marín-Riera's results reveal a lot about how evolution may work in practice. The originality of their work lies in how the model realistically maps the interactions between genetic parameters, developmental processes, adult phenotypes and functional properties. These interactions introduce dynamic properties that allow for morphological transitions and evolutionary novelties that were not captured by earlier computational models of the evolution of complex phenotypes¹⁰. The authors' results highlight interesting questions. If phenotypic evolution is likely to get stuck on suboptimal adaptive peaks, what happens to the evolving populations? Do they simply persist in a suboptimal state? Do their ecological relationships change to fit their phenotype, thereby creating new adaptive peaks? Or do they become extinct as they are out-competed by populations that can reach a more optimal phenotype? Salazar-Ciudad and Marín-Riera's focus on teeth — which can be studied empirically in living populations, among distantly related clades and in the fossil record — offers considerable potential for testing their evolutionary predictions and closing the knowledge gaps between genetics, development and macroevolution. ■

P. David Polly is in the Departments of Geological Sciences, Biology and Anthropology, Indiana University, Bloomington, Indiana 47405, USA. e-mail: pdpolly@indiana.edu

1. Van Valen, L. *J. Herpetol.* **8**, 109–121 (1974).
2. Salazar-Ciudad, I. & Marín-Riera, M. *Nature* **497**, 361–364 (2013).
3. Jernvall, J., Keränen, S. V. E. & Thesleff, I. *Proc. Natl Acad. Sci. USA* **97**, 14444–14448 (2000).
4. Tucker, A. & Sharpe, P. *Nature Rev. Genet.* **5**, 499–508 (2004).
5. Ungar, P. S. *Mammal Teeth: Origin, Evolution, and Diversity* (Johns Hopkins Univ. Press, 2000).
6. Salazar-Ciudad, I. & Jernvall, J. *Evol. Dev.* **6**, 6–16 (2004).
7. Wright, S. *Proc. 6th Annu. Congr. Genet.* **1**, 356–366 (1932).
8. Whitlock, M. C. & Phillips, P. C. *Trends Ecol. Evol.* **15**, 347–348 (2000).
9. Wainwright, M. et al. *Am. Nat.* **165**, E140–E154 (2005).
10. Polly, P. D. *Palaeontol. Electron.* **2**, 7.2.7A (2004).

Inferring ancient divergences requires genes with strong phylogenetic signals

Leonidas Salichos¹ & Antonis Rokas¹

To tackle incongruence, the topological conflict between different gene trees, phylogenomic studies couple concatenation with practices such as rogue taxon removal or the use of slowly evolving genes. Phylogenomic analysis of 1,070 orthologues from 23 yeast genomes identified 1,070 distinct gene trees, which were all incongruent with the phylogeny inferred from concatenation. Incongruence severity increased for shorter internodes located deeper in the phylogeny. Notably, whereas most practices had little or negative impact on the yeast phylogeny, the use of genes or internodes with high average internode support significantly improved the robustness of inference. We obtained similar results in analyses of vertebrate and metazoan phylogenomic data sets. These results question the exclusive reliance on concatenation and associated practices, and argue that selecting genes with strong phylogenetic signals and demonstrating the absence of significant incongruence are essential for accurately reconstructing ancient divergences.

Concatenation, the compilation and analysis of hundreds of genes as a single data set, has become the standard approach for determining major branches of the tree of life^{1–5}. However, incongruence stemming from either analytical errors in gene history reconstruction^{6,7} or the action of biological processes⁸, evidenced by disagreements between phylogenomic studies^{9–14}, argues that the histories of some lineages are better depicted by or more closely resemble networks of highly related trees¹⁵ and that concatenation might not be as robust as confidence indices indicate. To tackle incongruence, studies have adopted several practices, such as removing unstable taxa^{1–3}, that are useful but not always effective^{16–18}.

The *Saccharomyces* and *Candida* yeasts are excellent for examining phylogenomic practices in the presence of incongruence, owing to the presence of conflicting gene trees^{7,19} and the availability of two synteny databases^{20,21} for genome-wide identification of high-quality orthologues, minimizing the risk of incongruence from hidden paralogy^{22,23} and horizontal gene transfer²⁴. Importantly, levels of sequence divergence between yeasts are intermediate to those observed between vertebrates and animals, making them an appropriate model for the study of ancient divergences.

Analyses of 1,070 groups of orthologues (below we refer to groups of orthologues simply as genes) from 23 yeast genomes showed that although concatenation resolved the species phylogeny, several internodes of the extended majority-rule consensus (eMRC) phylogeny of the 1,070 underlying gene trees were weakly supported. None of the 1,070 gene trees agreed with each other, with the concatenation phylogeny or with the eMRC phylogeny. The novel measure we developed to quantify the observed incongruence showed that standard practices aimed at reducing incongruence had little impact. In agreement with current theoretical models^{9,16,25,26}, incongruence was more severe for shorter internodes that are deeper on the phylogeny. Notably, the selection of genes whose bootstrap consensus trees had high average clade support, or the selection of highly supported internodes, significantly reduced incongruence, arguing that inference in deep time depends critically on the identification of molecular markers with strong phylogenetic signals.

All gene trees differ from species phylogeny

We assembled a data set of 1,070 genes from 23 yeast genomes^{20,21,27} (Methods and Supplementary Table 1). Maximum-likelihood analysis

of the concatenation of all 1,070 genes produced a species phylogeny in which all 20 internodes exhibited 100% bootstrap support (Fig. 1a); we obtained identical results using Bayesian inference and one other type of maximum-likelihood software (Supplementary Fig. 1). Notably, all 1,070 gene trees were topologically distinct and none matched the topology inferred from concatenation analysis (Fig. 1b). However, the average tree distance between the 1,070 gene trees was much lower (normalized Robinson–Foulds²⁸ tree distance = 0.52; that is, two gene trees differed, on average, in 10.4 out of their 20 bipartitions) than that between randomly generated trees of the same taxon number (0.99; that is, two trees differed on average in 19.8 out of 20 bipartitions), indicating that the yeast gene trees have similar evolutionary histories.

Summarizing the 1,070 gene trees into an eMRC phylogeny produced a topology that was identical to the concatenation phylogeny (Fig. 1a). However, although 11 out of 20 internodes in the eMRC phylogeny had a gene-support frequency (GSF) of greater than 50%, 5 of the remaining 9 internodes had a GSF of less than 30% (Fig. 1a). Furthermore, the most prevalent conflicts to most of these weakly supported internodes had substantial GSF values (Supplementary Table 2). For example, the relative positions of *Candida glabrata*, *Saccharomyces castellii* and the *Saccharomyces* 'sensu stricto' clade suggest that there are five uniquely shared chromosomal rearrangements and a substantially higher number of uniquely shared gene losses between *C. glabrata* and *S. cerevisiae*, which indicates that divergence of *S. castellii* preceded that of *C. glabrata* from the *Saccharomyces* sensu stricto clade²². Although concatenation provided 100% bootstrap support for the apparently incorrect grouping of *S. castellii* with the *Saccharomyces* sensu stricto clade (Fig. 1a), only 311 out of 1,070 gene trees (29%) favoured it, whereas 214 (20%) favoured the grouping of *C. glabrata* with the *Saccharomyces* sensu stricto clade.

A novel measure that considers incongruence

To quantify incongruence, we developed 'internode certainty', which evaluates support for a given internode by considering its frequency in a given set of trees jointly with that of the most prevalent conflicting bipartition in the same set of trees. Like phylogenetic network methods developed for visualizing phylogenetic conflicts¹⁵, internode certainty relies on the bipartitions present in trees, each of which is a split of

¹Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee 37235, USA.

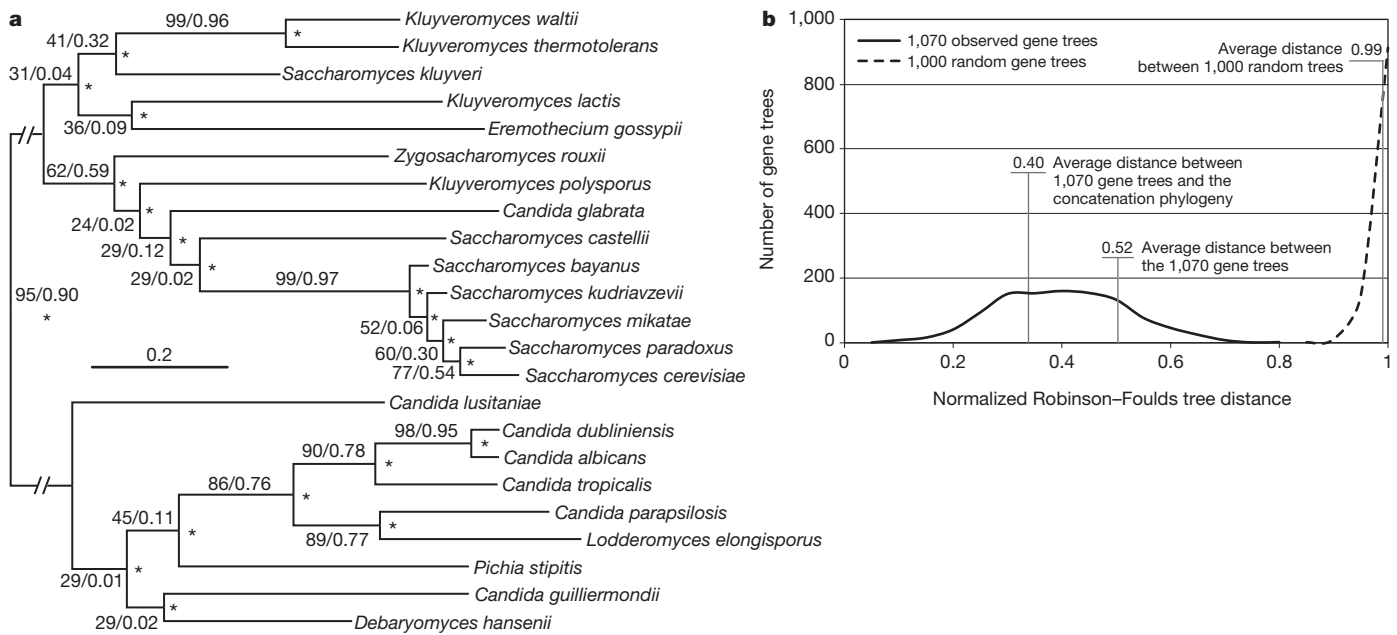


Figure 1 | The yeast species phylogeny recovered from the concatenation analysis of 1,070 genes disagrees with every gene tree, despite absolute bootstrap support. **a**, The yeast species phylogeny recovered from concatenation analysis of 1,070 genes using maximum likelihood. Asterisks denote internodes that received 100% bootstrap support by the concatenation analysis. Values near internodes correspond to gene-support frequency and internode certainty, respectively. The scale bar is in units of amino-acid substitutions per site. **b**, The distribution of the agreement between the bipartitions present in the 1,070 individual gene trees and the concatenation phylogeny, as well as the distribution of the agreement

the taxa into two mutually exclusive non-empty groups. Compared to other incongruence measures^{29–32}, internode certainty is not character-based^{29–31}, does not depend on an optimality criterion^{29–31} or clade support metric³², and can be applied to any set of trees. For example, if the entire set of gene trees is used, the internode certainty of a given internode will reflect the amount of information available for that internode in the set of gene trees by considering the internode's GSF jointly with the GSF of the most prevalent bipartition that conflicts with the internode. If the set of bootstrap replicate trees for a given gene is used, then internode certainty will be calculated based on bootstrap-support values. Internode-certainty values near zero indicate the presence of an almost equally supported bipartition that conflicts with the inferred internode, whereas values close to one indicate the absence of conflict. Examination of the eMRC phylogeny showed that 9 out of 20 internodes had an internode certainty of less than 0.3, which corresponds to a less than 4:1 ratio between the support for the inferred internode and its most prevalent conflicting bipartition, and that 7 out of 20 internodes had an internode certainty of less than 0.1 (a ratio of less than 7:3) (Fig. 1a and Supplementary Fig. 2).

Given that internode certainty measures the degree of conflict for every internode, it is more informative than GSF. For example, the placement of *Saccharomyces bayanus* and the placement of *Zygosaccharomyces rouxii* received 52% and 62% GSF, whereas their internode certainties were 0.06 and 0.59, respectively (Fig. 1a). This marked difference in the internode-certainty values of the two internodes (despite similar GSF values) is a result of the strong secondary conflicting signal in *S. bayanus* only³³ (29% GSF for grouping *S. bayanus* with *Saccharomyces kudriavzevii*), and not in *Z. rouxii* (Supplementary Table 2). Furthermore, comparison of the sums of internode-certainty values across trees of a given taxon number ('tree certainty') can be used to quantify changes in the degree of incongruence between trees inferred from different data sets or methods.

between the bipartitions present in 1,000 randomly generated trees of equal taxon number and the concatenation phylogeny, measured using the normalized Robinson–Foulds tree distance. Average distances between the 1,070 gene trees and the concatenation phylogeny, between the 1,070 gene trees themselves, and between 1,000 randomly generated gene trees that have equal taxon numbers, are also shown. The phylogeny of the 23 yeast species analysed in this study is unrooted and contains 20 non-trivial bipartitions; because the divergence of *Saccharomyces* and *Candida* lineages is well established, the mid-point rooting of the phylogeny is shown for easier visualization.

Standard practices do not reduce incongruence

To test whether we could decrease incongruence, we evaluated the effect of several standard phylogenomic practices purported to do so on the inference of the yeast phylogeny (Fig. 2). Specifically, we tested the effect of: removing sites containing gaps as well as 'rogue' genes that produced alignments of bad quality (Supplementary Fig. 3); removing unstable and quickly evolving species (Supplementary Figs 4–6); using only genes that recover a particular internode widely regarded as certain, or well established, from prior data (Supplementary Figs 6 and 7); using only slowly evolving genes (Supplementary Fig. 8); and using conserved amino acid substitutions or indels (Supplementary Fig. 9).

The first three practices did not have a substantial effect on the inference and support of the yeast phylogeny, whereas the use of slowly evolving genes and conserved sites increased incongruence across many internodes of the yeast phylogeny (Fig. 2). Furthermore, the removal of unstable or quickly evolving species from the *Saccharomyces* lineage had no effect on, often highly ambiguous, internodes in the *Candida* lineage and vice versa (Supplementary Figs 5 and 6), arguing that the impact of removing rogue taxa was not only minimal but also highly localized.

Support depends on internode length and depth

Examination of whether the degree of incongruence, as measured by low GSF, correlated with internode length and depth, as measured by branch lengths, showed that incongruence was stronger in early divergent and short internodes (Fig. 3). This is consistent with theoretical expectations^{9,16,25,26}. To test whether this relationship is the same in other lineages, we generated a data set of 1,086 genes from 18 vertebrate species, which has higher sequence similarity than the yeast data set (61% versus 44% average pairwise amino acid similarity, respectively), and a data set of 225 genes from metazoan species,

Treatment	Treatment details	Average GSF	Tree certainty	GSF increases	GSF decreases	IC increases	IC decreases	Average GSF	Tree certainty
Removal of sites containing gaps	Default analysis	60.02	8.35	-	-	-	-	48	5.18
	All sites with gaps are excluded	58.17	7.91	0	5	0	7	50	5.69
	All sites with $\geq 50\%$ gaps are excluded	60.04	8.23	0	0	1	2	52	6.20
Removal of poorly aligned genes	Default analysis ($x = 50\%$; 1,070 genes)	60.00	8.35	-	-	-	-	54	6.71
	Poor alignments removed ($x = 70\%$; 374 genes)	60.24	8.42	2	1	4	3	56	7.22
Removal of quickly evolving or unstable species	<i>C. lusitaniae</i> (unstable)	62.22	8.15	1	0	2	2	58	7.73
	<i>S. castellii</i> (unstable)	62.08	8.20	1	0	1	1	60	8.24
	<i>K. polysporus</i> (fast and unstable)	63.30	8.33	3	0	1	1	62	8.75
	<i>E. gossypii</i> (fast and unstable)	61.93	7.98	2	0	0	4	64	9.26
	<i>C. glabrata</i> (fast and unstable)	63.10	8.30	3	0	1	2	66	9.77
	<i>K. lactis</i> (fast and unstable)	61.86	7.99	2	1	0	3	68	10.28
	<i>E. gossypii</i> , <i>K. lactis</i>	63.91	7.88	1	1	0	3	70	10.79
	<i>E. gossypii</i> , <i>C. glabrata</i> , <i>K. lactis</i>	67.32	7.88	3	0	1	3	72	11.30
Selection of genes that recover specific bipartitions	(<i>C. glabrata</i> , <i>S. bayanus</i> , <i>S. kudriavzevii</i> , <i>S. mikatae</i> , <i>S. cerevisiae</i> , <i>S. paradoxus</i>)	65.88	9.47	4	1	6	3		
	(<i>Z. rouxii</i> , <i>K. polysporus</i> , <i>C. glabrata</i> , <i>S. bayanus</i> , <i>S. castellii</i> , <i>S. kudriavzevii</i> , <i>S. mikatae</i> , <i>S. cerevisiae</i> , <i>S. paradoxus</i>)	63.34	8.62	3	0	0	4		
	(<i>C. tropicalis</i> , <i>C. dubliniensis</i> , <i>C. albicans</i>)	61.20	8.62	1	0	0	0		
Selection of the most slowly evolving genes	The 100 slowest evolving genes	52.20	6.76	1	10	2	9		
Selection of genes whose bootstrap consensus trees have high average BS	Genes with average BS $\geq 60\%$ (904 genes)	62.17	8.59	4	0	2	0		
	Genes with average BS $\geq 70\%$ (545 genes)	65.68	9.18	14	0	12	0		
	Genes with average BS $\geq 80\%$ (131 genes)	70.56	9.92	15	0	14	0		
Selection of genes whose bootstrap consensus trees have high tree certainty	Using only the 904 genes with the highest TC	62.26	8.72	6	0	2	0		
	Using only the 545 genes with the highest TC	66.06	9.37	13	0	12	0		
	Using only the 131 genes with the highest TC	71.20	10.28	16	0	12	1		
Selection of bipartitions with high BS in the bootstrap consensus trees of genes	Using only bipartitions that have $\geq 60\%$ BS	NA	10.11	-	-	14	0		
	Using only bipartitions that have $\geq 70\%$ BS	NA	10.70	-	-	16	0		
	Using only bipartitions that have $\geq 80\%$ BS	NA	11.32	-	-	15	0		

Figure 2 | Differences in yeast phylogenies inferred from different phylogenomic practices. The specific phylogenomic practice tested (Treatment), the average GSF of the internodes of the yeast phylogeny, the tree certainty (TC) of the yeast phylogeny, the numbers of internodes of the yeast phylogeny in which GSF increases or decreases by more than 3% (GSF increases and GSF decreases), and the numbers of internodes of the yeast phylogeny in which internode certainty increases or decreases by more than 0.03 (internode certainty (IC) increases and IC decreases). As the maximum value of internode certainty for a given internode is 1, the maximum value of tree certainty for a

given phylogeny is the number of internodes, which will equal $K - 3$, where K is the number of taxa used. In the analyses concerned with the removal of poorly aligned genes, only genes whose alignment length after gap removal is greater than or equal to a certain percentage, x , of the original alignment were used. In the analyses concerned with the use of bipartitions, only those bipartitions that displayed bootstrap support greater or equal to 60%, 70% or 80% in the bootstrap consensus trees of the 1,070 genes were used to construct eMRC phylogenies, which were then compared with the default analysis. NA, not applicable.

which has lower sequence similarity (29% average pairwise amino acid similarity). The vertebrate genes produced 299 distinct gene trees (average normalized Robinson–Foulds tree distance = 0.42). Concatenation analysis suggested a completely supported species phylogeny; however, this phylogeny was topologically identical to 15 gene trees and eMRC analysis showed that 4 out of 15 internodes had a GSF of less than 50% and internode certainty of less than 0.3 (Supplementary Fig. 10a–c). Similarly, the 225 metazoan genes produced 224 distinct gene trees (average normalized Robinson–Foulds tree distance = 0.72). Concatenation analysis suggested 14 out of 18 internodes with 100% bootstrap support, despite the fact that it was not topologically identical to any of the 225 gene trees and that 10 out of 18 internodes had less than 50% GSF and less than 0.1 internode certainty (Supplementary Fig. 10d–f). Interestingly, incongruence was significantly correlated only with short internodes in the (less divergent) vertebrates, nearly equally significantly with both internode length and internode depth in yeasts, and more significantly with internode depth than with internode length in the (more divergent) metazoans (Fig. 3).

Strong signal reduces incongruence

To test whether the selection of genes with stronger phylogenetic signal reduced incongruence, we analysed three data sets comprising genes whose bootstrap consensus trees showed average bootstrap support across all internodes that was greater than or equal to 60% (904 genes), 70% (545 genes), or 80% (131 genes), and three data sets comprising the 904, 545 or 131 genes whose bootstrap consensus trees had the highest tree certainty. Selecting genes with high average bootstrap support or high tree certainty significantly reduced incongruence across many, but not all, internodes (Fig. 2, and Supplementary Figs 11 and

12). Concatenation analysis of the sets of genes with average bootstrap support that was greater than or equal to 60% and 70% (and of the 904 and the 545 genes with the highest tree certainty) produced the same species phylogeny as when all genes were analysed. Notably, analysis of genes with average bootstrap support that was greater than or equal to 80%, as well as of the 131 genes with the highest tree certainty, produced the correct placement of *C. glabrata* (Supplementary Fig. 11c, f). This result, to our knowledge, has not been observed in any concatenation-based yeast phylogenomic analysis^{7,34–37}, and suggests that high bootstrap support is a good indicator of a gene's phylogenetic usefulness, but also that concatenating genes with high bootstrap support reduces incongruence and improves resolution.

We also tested whether selecting internodes with high bootstrap support decreased incongruence by extracting only those bipartitions that displayed bootstrap support values $\geq 60\%$, $\geq 70\%$, and $\geq 80\%$ from every one of the 1,070 genes' bootstrap consensus trees and then using them to construct new eMRC phylogenies (Supplementary Figs 12 and 13). One advantage of working with taxon bipartitions, rather than genes, is that we can quantify a given internode's internode certainty from only the subset of bipartitions that highly support or conflict with that internode. This practice significantly increased internode certainty values for ≥ 14 internodes relative to the phylogeny of Fig. 1a and showed the highest tree certainty of all our analyses (Fig. 2). Interestingly, while internode certainty for most internodes increased when we increased the bootstrap support threshold, this was not the case for several of the most difficult to resolve internodes (Supplementary Fig. 13d), suggesting that those few genes that show high bootstrap support for short internodes deep in the phylogeny strongly conflict with each other. We obtained similar results when we

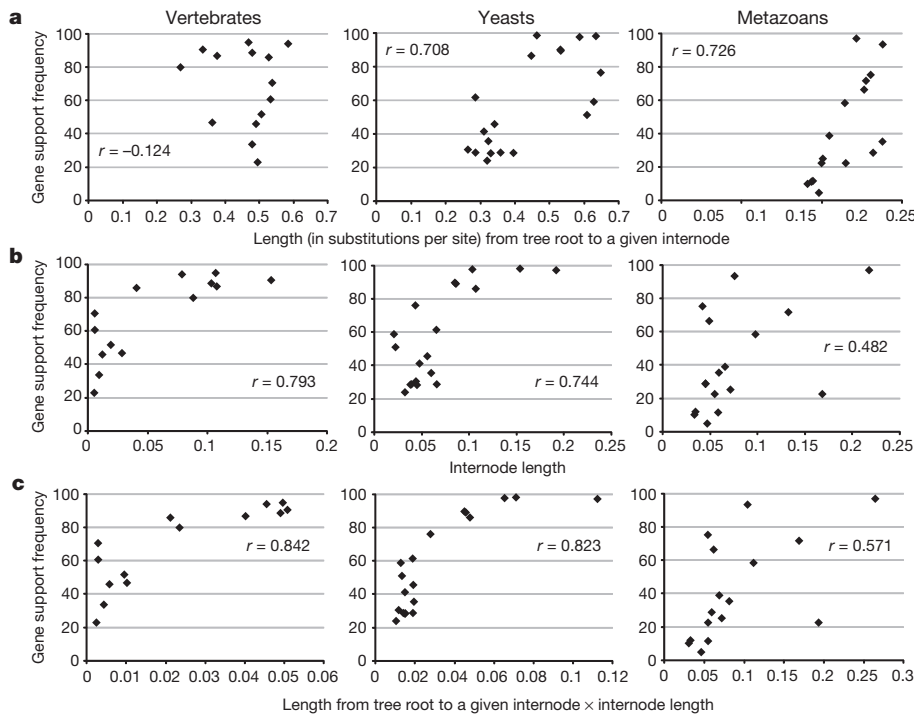


Figure 3 | Incongruence is more prevalent in shorter internodes located deeper on the phylogeny. The correlation (Pearson's r) between a measure of internode support (GSF) with internode length and depth was measured for each internode present in three data sets that show lower (vertebrates, 1,086 genes), intermediate (yeasts, 1,070 genes) and higher (metazoans, 225 genes) levels of sequence divergence. **a**, GSF is positively correlated with internode length in yeasts and metazoans. **b**, GSF is positively correlated with the root to internode length in all three lineages, indicating that internodes placed deeper in the phylogeny typically have lower GSF. **c**, GSF is positively correlated with the product of internode length and root to internode length in all three lineages.

performed the same analyses on the vertebrate and metazoan data sets (Supplementary Fig. 14).

Standard practices can mislead

We constructed and analysed 1,070 yeast genes with the aim of determining the yeast phylogeny. If only concatenation and standard phylogenomic practices had been used, this analysis would have produced an absolutely supported phylogeny similar to those obtained by major phylogenomic studies^{1,3–5,11,12,16,19}. However, examination of the signal in gene trees showed that concatenation masked the considerable incongruence present in several internodes. Thus, while analyses of approximately 20% of the genes typically present in a yeast genome definitively support many internodes of the yeast phylogeny, the topology of a considerable number of others remains uncertain (Supplementary Figs 15 and 16).

Our finding that incongruence correlates with early divergent and short internodes indicates that analytical factors are major contributors; however, it is likely that biological factors have also contributed. 'Species tree' methods use coalescent theory to estimate the species phylogeny from the individual gene trees, allowing for lineage sorting, a common biological explanation for gene trees that are incongruent with the species phylogeny⁸. Unfortunately, many such methods assume that analytical errors in inference are minimal, a valid assumption for most shallow clades but one that is untenable for the deeply divergent clades of the yeast phylogeny. For example, analysis of our data set with the average unit-ranking method³⁸ produced a species phylogeny in which all the internodes with very low GSF and internode-certainty values were extremely short, mainly because all incongruence was considered to be due to variation in coalescent depth across gene trees (Supplementary Fig. 17a). Not surprisingly, these coalescent unit-based branch lengths were highly correlated with internodes' GSF and internode-certainty values (Supplementary Fig. 17b). Furthermore, bootstrapping of this data set produced a highly supported species phylogeny (Supplementary Fig. 17a), again contradicting our findings of extensive conflict in certain internodes.

Perspective

These results argue that elimination of the observed incongruence between phylogenomic studies^{1,3,4,11,12} will require three fundamental

revisions to current practices. First, we should abandon using bootstrap support on concatenation analyses of large data sets. Bootstrapping was developed long before the discovery of high-throughput sequencing, and it is an extremely useful measure of sampling error—that is, the robustness of inference when data are limited³⁹—such as when a single gene is analysed. Given the availability and ease of generating genome-scale data⁴⁰, relying on bootstrap to analyse phylogenomic data sets is misleading, not only because sampling error is minimal but also because its application will, even in the presence of notable conflict⁹ or systematic error^{6,16}, almost always result in 100% values^{9,19,41}.

The second critical revision is that we carefully examine the signal present in individual genes^{16,29–32,42} and their trees¹⁵. Our results indicate that the subset of genes with strong phylogenetic signal is more informative than the full set of genes, suggesting that phylogenomic analyses using conditional combination approaches, rather than approaches based on total evidence, may be more powerful⁴². Preferably, such analyses would be combined with internode-specific approaches³¹ because the latter can uncover internodes that harbour several conflicting phylogenetic signals. As the internode certainty measure shows (Supplementary Fig. 2), the amount of information for a given internode that is supported by 50% of gene trees, with the remaining 50% being uninformative, is far greater than in cases in which the remaining 50% of the gene trees support two or three alternative conflicting topologies. In the first case the gene trees strongly suggest that the internode is resolved, whereas in the second there is reason to be cautious.

Finally, it is necessary to identify explicitly internodes that, despite the use of genome-scale data sets, robust study designs and powerful algorithms, are poorly supported. We argue that the ongoing debate regarding phylogenies inferred from different phylogenomic studies¹⁰ concerns internodes that are poorly supported by individual gene trees. Identifying these internodes and distinguishing them from those that are supported by a considerable fraction of genes and that lack conflicts will be far more beneficial than simply helping to pinpoint challenging internodes. It should enable us to identify the broad contours of the network of highly related gene histories that is the tree of life. Perhaps most importantly, it will focus the attention of researchers to develop novel phylogenomic approaches and markers to more accurately decipher the most challenging ancient branches of life's genealogy from the DNA record.

METHODS SUMMARY

Using synteny and orthology information present in the YGOB²⁰ and CGOB²¹ databases from 23 yeast genomes^{20,21,27}, we constructed an initial data set of 2,651 genes that, following quality control (see Methods), was reduced to the final 1,070. We also used the complete gene sets from 18 vertebrate and 21 metazoan species and used the cRBH algorithm²³ to identify 1,086 vertebrate and 225 metazoan genes. Genes were aligned using MAFFT⁴³, the best-fit evolutionary model was determined using ProtTest⁴⁴, and the maximum-likelihood tree was estimated using RAxML⁴⁵. eMRC trees were determined using the Phylogeny Inference Package (PHYLIP; J. Felsenstein, University of Washington, Seattle; <http://evolution.genetics.washington.edu/phylip.html>) and custom Perl scripts. A series of different data sets was constructed using custom Perl scripts. Internode certainty, our novel measure that evaluates support for a given internode by considering its GSF (or bootstrap support) jointly with that of the most prevalent conflicting bipartition in the entire set of gene trees (or bootstrap replicate trees), was calculated according to the equation:

$$\text{Internode certainty} = \log_2(2) + p(x_1/(x_1 + x_2)) \log_2(p(x_1/(x_1 + x_2))) + p(x_2/(x_1 + x_2)) \log_2(p(x_2/(x_1 + x_2)))$$

where x_1 and x_2 are the frequencies of the first and second most prevalent conflicting bipartitions for a given internode.

Full Methods and any associated references are available in the online version of the paper.

Received 6 December 2012; accepted 28 March 2013.

Published online 8 May 2013.

- Dunn, C. W. *et al.* Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**, 745–749 (2008).
- Rokas, A., Kruger, D. & Carroll, S. B. Animal evolution and the molecular signature of radiations compressed in time. *Science* **310**, 1933–1938 (2005).
- Philippe, H. *et al.* Phylogenomics revisits traditional views on deep animal relationships. *Curr. Biol.* **19**, 706–712 (2009).
- Schierwater, B. *et al.* Concatenated analysis sheds light on early metazoan evolution and fuels a modern “urmetazoan” hypothesis. *PLoS Biol.* **7**, e20 (2009).
- Regier, J. C. *et al.* Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* **463**, 1079–1083 (2010).
- Phillips, M. J., Delsuc, F. D. & Penny, D. Genome-scale phylogeny and the detection of systematic biases. *Mol. Biol. Evol.* **21**, 1455–1458 (2004).
- Hess, J. & Goldman, N. Addressing inter-gene heterogeneity in maximum likelihood phylogenomic analysis: yeasts revisited. *PLoS ONE* **6**, e22783 (2011).
- Degnan, J. H. & Rosenberg, N. A. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* **24**, 332–340 (2009).
- Rokas, A. & Carroll, S. B. Bushes in the tree of life. *PLoS Biol.* **4**, e352 (2006).
- Philippe, H. *et al.* Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* **9**, e1000602 (2011).
- Kocot, K. M. *et al.* Phylogenomics reveals deep molluscan relationships. *Nature* **477**, 452–456 (2011).
- Smith, S. J. *et al.* Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature* **480**, 364–367 (2011).
- Bourliat, S. J. *et al.* Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* **444**, 85–88 (2006).
- Delsuc, F., Brinkmann, H., Chourrout, D. & Philippe, H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**, 965–968 (2006).
- Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
- Regier, J. C. *et al.* Resolving arthropod phylogeny: exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. *Syst. Biol.* **57**, 920–938 (2008).
- Regier, J. C. & Zwick, A. Sources of signal in 62 protein-coding nuclear genes for higher-level phylogenetics of arthropods. *PLoS ONE* **6**, e23408 (2011).
- Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).
- Rokas, A., Williams, B. L., King, N. & Carroll, S. B. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**, 798–804 (2003).
- Byrne, K. P. & Wolfe, K. H. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res.* **15**, 1456–1461 (2005).
- Fitzpatrick, D. A., O’Gaora, P., Byrne, K. P. & Butler, G. Analysis of gene evolution and metabolic pathways using the *Candida* Gene Order Browser. *BMC Genomics* **11**, 290 (2010).
- Scannell, D. R., Byrne, K. P., Gordon, J. L., Wong, S. & Wolfe, K. H. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* **440**, 341–345 (2006).
- Salichos, L. & Rokas, A. Evaluating ortholog prediction algorithms in a yeast model clade. *PLoS ONE* **6**, e18755 (2011).
- Slot, J. C. & Rokas, A. Multiple GAL pathway gene clusters evolved independently and by different mechanisms in fungi. *Proc. Natl Acad. Sci. USA* **107**, 10136–10141 (2010).
- Mossel, E. & Steel, M. A phase transition for a random cluster model on phylogenetic trees. *Math. Biosci.* **187**, 189–203 (2004).
- Townsend, J. P., Su, Z. & Tekle, Y. I. Phylogenetic signal and noise: predicting the power of a data set to resolve phylogeny. *Syst. Biol.* **61**, 835–849 (2012).
- Scannell, D. R. *et al.* The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. *G3* **1**, 11–25 (2011).
- Robinson, D. R. & Foulds, L. R. Comparison of phylogenetic trees. *Math. Biosci.* **53**, 131–147 (1981).
- Farris, J. S., Källersjö, M., Kluge, A. G. & Bult, C. Testing significance of incongruence. *Cladistics* **10**, 315–319 (1995).
- Templeton, A. R. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and apes. *Evolution* **37**, 221–244 (1983).
- Baker, R. H. & DeSalle, R. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. *Syst. Biol.* **46**, 654–673 (1997).
- Rodrigo, A. G., Kelly-Borges, M., Bergquist, P. G. & Bergquist, P. L. A randomisation test of the null hypothesis that two cladograms are sample estimates of a parametric phylogenetic tree. *N. Z. J. Bot.* **31**, 257–268 (1993).
- Yu, Y., Degnan, J. H. & Nakhleh, L. The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLoS Genet.* **8**, e1002660 (2012).
- Hittinger, C. T., Rokas, A. & Carroll, S. B. Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proc. Natl Acad. Sci. USA* **101**, 14144–14149 (2004).
- Rokas, A. & Carroll, S. B. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Mol. Biol. Evol.* **22**, 1337–1344 (2005).
- Jeffroy, O., Brinkmann, H., Delsuc, F. & Philippe, H. Phylogenomics: the beginning of incongruence? *Trends Genet.* **22**, 225–231 (2006).
- Fitzpatrick, D. A., Logue, M. E., Stajich, J. E. & Butler, G. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evol. Biol.* **6**, 99 (2006).
- Liu, L., Yu, L., Pearl, D. K. & Edwards, S. V. Estimating species phylogenies using coalescence times among sequences. *Syst. Biol.* **58**, 468–477 (2009).
- Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791 (1985).
- Hittinger, C. T., Johnston, M., Tossberg, J. T. & Rokas, A. Leveraging skewed transcript abundance by RNA-seq to increase the genomic depth of the tree of life. *Proc. Natl Acad. Sci. USA* **107**, 1476–1481 (2010).
- Kumar, S., Filipski, A. J., Battistuzzi, F. U., Kosakovsky Pond, S. L. & Tamura, K. Statistics and truth in phylogenomics. *Mol. Biol. Evol.* **29**, 457–472 (2012).
- Cunningham, C. W. Can three incongruence tests predict when data should be combined? *Mol. Biol. Evol.* **14**, 733–740 (1997).
- Katoh, K. & Toh, H. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* **9**, 286–298 (2008).
- Abascal, F., Zardoya, R. & Posada, D. Prottest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104–2105 (2005).
- Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank K. Polzin for providing a script that identified alignment sites that contained single substitutions between amino acids that differ in their physicochemical properties. We thank members of the Rokas laboratory and B. O’Meara for valuable comments on this work. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University. This work was supported by the National Science Foundation (DEB-0844968).

Author Contributions L.S. and A.R. conceived and designed experiments; L.S. carried out experiments; L.S. and A.R. analysed data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.R. (antonis.rokas@vanderbilt.edu).

METHODS

Data matrix construction. We used the complete sets of annotated genes from 23 yeast genomes^{20,21,27,46} (Supplementary Table 1) and, using the synteny and orthology information present in the YGOB²⁰ and CGOB²¹ databases, we constructed an initial data set of 2,651 groups of orthologues (referred to below simply as genes) that had representatives in all 23 genomes. This reliance on two highly accurate and manually curated synteny databases and the requirement for a given orthologue to be present in all 23 species greatly minimized errors in orthology inference due to hidden paralogy^{23,47}. It also avoided the inclusion of any horizontally transferred genes present in some, but not all, species as well as any horizontally transferred genes present in regions that lack synteny conservation. For any potentially horizontally transferred gene to be included in our data matrix, it would have had to have been gained in some, but not all, yeast species used in our study and it would have had to replace the native gene and take up its position on the chromosome, which has never been observed in yeasts^{24,48–50} and is probably very rare.

The nucleotide sequences of all genes were translated to amino acids, taking into account that in certain species in the *Candida* lineage the CUG codon encodes for the amino acid serine rather than leucine. Using alignment quality and individual-gene-length filtering criteria described below, we then reduced the number of genes to 1,070. Examination of the functional annotation—as defined by the Gene Ontology consortium⁵¹—of the 1,070 *S. cerevisiae* orthologues using the Gostat software⁵² showed that this gene set is statistically overrepresented for several different functional categories, such as cellular metabolic process, cellular component organization and biogenesis, and ribosome assembly and biogenesis; that is, for categories associated with standard cell housekeeping functions. Analysis of different orthologue subsets (for example, of the 131 genes whose bootstrap consensus trees show the highest average bootstrap support) shows that for many of these subsets, some of the same functions were statistically overrepresented.

We also created two additional data sets from the complete sets of annotated genes from 18 vertebrate and 21 metazoan species (Supplementary Table 1). The two data sets were constructed using the cRBH algorithm²³, and were comprised of 1,086 vertebrate and 225 metazoan genes. To avoid constructing groups of orthologues that contained very distant homologues, we set the filtering parameter of the cRBH algorithm²³, which considers the degree by which the two proteins differed in sequence length or BLAST alignment, to $r = 0.3$.

For each species, for reasons of space and convenience, we constructed a corresponding acronym using the first letter from the genus name and the three first letters from the species name (for example, the acronym for *S. cerevisiae* is 'Scer'; see Supplementary Information). All data matrices are available from the authors on request.

Gene alignment and filtering criteria. To minimize the use of genes that contained sequences whose annotation was problematic, or which resulted in alignments of low quality, we applied various filtering criteria. We first excluded, before alignment, all genes that had an average sequence length of less than or equal to 150 amino acids. Second, we aligned all genes using the MAFFT software⁴³, using the default settings, and excluded genes whose alignments after removing all positions that contained gaps were less than or equal to 50% of the original alignment length.

Gene-tree inference. For each gene, the best-fit evolutionary model was selected using ProtTest⁴⁴. The models typically consisted of an empirically determined amino acid substitution matrix (for example, WAG⁵³), empirically measured amino acid state frequencies and accounted for heterogeneity in evolutionary rates among sites by using the gamma distribution as well as by allowing for a given proportion of sites to be invariable. The unrooted phylogenetic tree of each and every gene, also called the gene tree, was then inferred using RAxML⁴⁵.

Species phylogeny inferred from concatenation and eMRC approaches. For the concatenation analysis, gene alignments were analysed as a single supermatrix. An unrooted concatenation species phylogeny was then inferred under the 'PROTGAMMAIWAG' model of amino acid substitution in RAxML⁴⁵, and confirmed with GARLI⁵⁴ as well as with MrBayes⁵⁵. The unrooted eMRC phylogeny that consisted of those bipartitions that appear in more than half of the maximum-likelihood estimated gene trees, as well as of additional compatible bipartitions that appear in less than half of the gene trees^{56,57}, was inferred from the CONSENSE program in PHYLIP. The eMRC phylogeny of bipartitions with high bootstrap support was constructed using custom Perl scripts. As the divergence of *Saccharomyces* and *Candida* lineages is well established, all phylogenies shown in figures have been mid-point rooted at the internode that separates these two lineages for easier visualization.

Species phylogeny inference using a consensus phylogenetic network approach. A consensus phylogenetic network was constructed based on the 1,070 gene trees

estimated by maximum likelihood using the median network construction algorithm in the SplitsTree4 software¹⁵ with a threshold of 0.1.

Tree distance estimation. Distances between trees were estimated using the normalized Robinson–Foulds tree distance²⁸, as calculated by RAxML⁴⁵. Sets of random trees for 23 taxa (yeasts), 18 taxa (vertebrates), and 21 taxa (metazoans), were generated using the random tree generator in the T-REX webserver³⁸, using the random tree generation procedure described by Kuhner and Felsenstein⁵⁹.

Internode certainty. A phylogenetic tree is an acyclic connected graph that represents evolutionary relationships among different genes or taxa and consists of nodes that are connected by edges or internodes. Phylogenetic trees can also be represented in a variety of other ways. One useful depiction is as sets of bipartitions (or splits). In this representation, each internode in a phylogenetic tree is viewed as a bipartition between two sets of taxa. For example, given a set of five species (*S. cerevisiae*, *Saccharomyces paradoxus*, *Saccharomyces mikatae*, *S. kudriavzevii* and *S. bayanus*), one example of a bipartition is one that separates the set of *S. cerevisiae*, *S. paradoxus* and *S. mikatae* from the set of *S. kudriavzevii* and *S. bayanus*.

Information from multiple phylogenetic trees from the same set of taxa is typically summarized using consensus trees. For example, the MRC approach⁵⁶ calculates the shared bipartitions across all phylogenetic trees and displays only those shared by their majority. Consequently, each internode in the MRC tree typically contains a value that corresponds to the percentage of individual trees that contain a given bipartition, but does not provide any information about the next most prevalent conflicting bipartition, or more generally, about the distribution of bipartitions that conflict with the internode. For example, if a consensus tree reports that 51 out of 100 phylogenetic trees contain a specific bipartition, we are not informed about whether the second-most prevalent conflicting bipartition is present in the remaining 49 trees or in 5 of the remaining trees. However, the first case (51% versus 49%) would indicate that both bipartitions have nearly equal support, whereas the second case (51% versus 5%) would indicate that the first bipartition is the only strongly supported bipartition for this internode. Consensus phylogenetic networks^{15,60}, which are potentially hyperdimensional graphs inferred from all bipartitions present above a certain frequency in a given set of trees, are very useful in visualizing such conflicting bipartitions. To quantify the degree of incongruence, as well as examine whether incongruence is reduced when standard phylogenomic practices are applied, we developed internode certainty, a measure that provides robust quantitative measures of the information conveyed by conflicting bipartitions for each internode.

Shannon's entropy measures the amount of certainty found in a random variable⁶¹. For example, when tossing a fair coin, heads or tails are equally probable and so the amount of certainty that we have about the outcome is 0, whereas if the coin is not fair, our certainty about the toss outcome will be high. Similarly, we can quantify the certainty that we have in the deduction of a given internode in a phylogenetic tree, by introducing a function that is maximized in the absence of any conflicting bipartitions but is minimized in the presence of equally prevalent conflicting bipartitions. Internode certainty quantifies the certainty of a bipartition that appears on a phylogenetic tree (that is, of a given internode) by considering its frequency of occurrence against that of the second most prevalent conflicting bipartition. Specifically, for the two most prevalent conflicting bipartitions:

$$\text{Internode certainty} = \log_2(2) + p(x_1/(x_1 + x_2)) \log_2(p(x_1/(x_1 + x_2))) \\ + p(x_2/(x_1 + x_2)) \log_2(p(x_2/(x_1 + x_2)))$$

where x_1 and x_2 are the frequencies of the first and second most prevalent conflicting bipartitions for a given internode.

Internode certainty, as well as the related measure called tree certainty (see below), can be measured on any given set of trees. For example, if the entire set of gene trees is used, the internode certainty value of a given internode will reflect the amount of information available for that internode in the set of gene trees by considering the internode's GSF jointly with the GSF of the most prevalent bipartition that conflicts with the internode. If the set of bootstrap replicate trees for a given gene is used, then internode certainty will be calculated based on bootstrap support values (instead of GSF values). Internode certainty can also be measured on any given set of bipartitions. For example, any two-state character that is variable across x species can be thought of as a bipartition, as it splits the set of taxa into two distinct groups. Thus, one can use internode certainty to measure the amount of information available for a given bipartition, and quantify the extent of incongruence, by considering the number of characters supporting that bipartition jointly with the number of characters supporting the most prevalent bipartition that conflicts with the internode. For example, if we assume that there

are four prevalent conflicting bipartitions with frequencies of 40%, 10%, 10% and 10%, respectively, for a given internode, then:

$$\text{Internode certainty} = 1 + (40/(40 + 10)) \log_2(40/(40 + 10)) + (10/(40 + 10)) \log_2(10/(40 + 10)) \approx 0.28$$

As another example, if we assume that there are four prevalent conflicting bipartitions with frequencies of 40%, 40%, 10% and 10%, respectively, for a given internode, then:

$$\text{Internode certainty} = 1 + (40/(40 + 40)) \log_2(40/(40 + 40)) + (40/(40 + 40)) \log_2(40/(40 + 40)) = 0.00$$

We define tree certainty as the sum of all internode-certainty values across all internodes of a phylogenetic tree.

Evaluation of phylogenomic practices. To remove positions or genes with gaps, we used custom Perl scripts to modify our default alignments by removing sites that contained greater than or equal to 50% gaps, or sites that contained any gap. We also tested whether the removal of genes producing alignments of bad quality improved inference of the species phylogeny by filtering genes whose alignment length after removal of all gap-containing sites was less than or equal to 70% of the original alignment length (instead of the less than or equal to 50% threshold used in the default analysis).

We removed several different unstable and quickly evolving species from the default data set, singly and in combination. After each removal, the gene sequences were re-aligned, a new best-fit evolutionary model was identified, and the phylogenetic analysis was carried out again with the new alignment and model.

To select genes that recover specific bipartitions, for the 100 bootstrap replicate trees constructed from each gene, we used the CONSENSE program in the PHYLIP package to generate the bootstrap consensus tree as well as its bipartitions. Using custom Perl scripts, we then extracted all genes that supported the three following bipartitions: first, *Candida albicans*, *Candida dubliniensis* and *Candida tropicalis*; second, *C. glabrata*, *Kluyveromyces polysporus*, *S. bayanus*, *S. castellii*, *S. cerevisiae*, *S. kudriavzevii*, *S. mikatae*, *S. paradoxus* and *Z. rouxii*; and third, *C. glabrata*, *S. bayanus*, *S. cerevisiae*, *S. kudriavzevii*, *S. mikatae* and *S. paradoxus*. We then used the selected genes and their gene trees to infer a species phylogeny using concatenation and eMRC analysis.

The 100 most slowly evolving genes were identified by calculating the 100 genes whose gene trees had the smallest sum of branch lengths.

To reduce the effect of homoplasy for early divergent internodes, many studies have suggested the use of rare substitution types⁶² as well as insertions or deletions (indels)⁶³. We constructed three data sets by extracting all sites from our 1,070 gene alignments that contained first, a single radical amino acid substitution (defined as a substitution with a blosum62 matrix score of less than or equal to -3) (20,289 sites); second, a single substitution between amino acids that differ radically in their physicochemical properties⁶² (4,075 sites); or third, a single indel that spans seven or more amino acids (2,474 sites). The presence of any of these three types of sites instantly parts a set of x species into two groups of taxa or, equivalently, into two bipartitions ($0_1 \dots 0_m$ and $1_1 \dots 1_n$), where $m \geq 2$ species contain the '0' character state, $n \geq 2$ species contain the '1' character state, and $m + n = x$. To quantify the extent of incongruence of each type of site on a given internode, we used internode certainty to measure the amount of information available for that internode by considering the number of characters supporting that internode jointly with the number of characters supporting the most prevalent bipartition that conflicts with the internode.

For every gene from the default data set, we estimated the average bootstrap support value of all 20 internodes of its bootstrap consensus tree. We also used the set of bootstrap replicate trees for every gene to calculate the internode-certainty value of every internode in its bootstrap consensus tree. When the internode certainty is calculated in this way, its value reflects the amount of information available for that internode in the set of bootstrap replicate trees because it considers the internode's bootstrap support jointly with the bootstrap support of the most prevalent bipartition that conflicts with the internode. We then calculated the tree-certainty value for each gene by summing the internode-certainty values of all internodes in its bootstrap consensus tree. Finally, we used these average bootstrap support and tree-certainty values to construct six subsets of orthogroups: three with genes that have average bootstrap support values that are greater than or equal to 60% (904 genes), 70% (545 genes) and 80% (131 genes), as well as three data sets of the 904, 545 and 131 genes with the highest tree certainty.

For every gene from the default data set, we extracted all bipartitions from its bootstrap consensus tree that had bootstrap support that was greater than or equal to 60%, 70% and 80%. We then used each one of these three sets of highly supported bipartitions to construct eMRC species phylogenies with custom Perl scripts.

We calculated the root-to-node length as the sum of all branch lengths from the midpoint of the rooted concatenation species phylogeny to the focal node. For the internode length, we considered the branch length of the internode leading to the focal node.

46. Dujon, B. Yeast evolutionary genomics. *Nature Rev. Genet.* **11**, 512–524 (2010).
47. Scannell, D. R., Butler, G. & Wolfe, K. H. Yeast genome evolution—the origin of the species. *Yeast* **24**, 929–942 (2007).
48. Hall, C., Brachat, S. & Dietrich, F. S. Contribution of horizontal gene transfer to the evolution of *Saccharomyces cerevisiae*. *Eukaryot. Cell* **4**, 1102–1115 (2005).
49. League, G. P., Slot, J. C. & Rokas, A. The ASP3 locus in *Saccharomyces cerevisiae* originated by horizontal gene transfer from *Wickerhamomyces*. *FEMS Yeast Res.* **12**, 859–863 (2012).
50. Novo, M. *et al.* Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proc. Natl Acad. Sci. USA* **106**, 16333–16338 (2009).
51. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nature Genet.* **25**, 25–29 (2000).
52. Beissbarth, T. & Speed, T. P. Gostat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **20**, 1464–1465 (2004).
53. Whelan, S. & Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691–699 (2001).
54. Zwickl, D. J. *Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets under the Maximum Likelihood Criterion*. Ph.D. thesis, Univ. Texas at Austin (2006).
55. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
56. Bryant, D. in *Bioconsensus* (eds M. Janowitz *et al.*) 163–184 (American Mathematical Society and DIMACS, 2003).
57. Felsenstein, J. *Inferring Phylogenies*. (Sinauer, 2003).
58. Alix, B., Boubacar, D. A. & Vladimirov, M. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res.* **40**, W573–W579 (2012).
59. Kuhner, M. K. & Felsenstein, J. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.* **11**, 459–468 (1994).
60. Holland, B. R., Huber, K. T., Moulton, V. & Lockhart, P. J. Using consensus networks to visualize contradictory evidence for species phylogeny. *Mol. Biol. Evol.* **21**, 1459–1461 (2004).
61. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
62. Rogozin, I. B., Wolf, Y. I., Carmel, L. & Koonin, E. V. Ecdysozoan clade rejected by genome-wide analysis of rare amino acid replacements. *Mol. Biol. Evol.* **24**, 1080–1090 (2007).
63. Belinky, F., Cohen, O. & Huchon, D. Large-scale parsimony analysis of metazoan indels in protein-coding genes. *Mol. Biol. Evol.* **27**, 441–451 (2010).

Structural and molecular interrogation of intact biological systems

Kwanghun Chung^{1,2}, Jenelle Wallace¹, Sung-Yon Kim¹, Sandhiya Kalyanasundaram², Aaron S. Andalman^{1,2}, Thomas J. Davidson^{1,2}, Julie J. Mirzabekov¹, Kelly A. Zalocusky^{1,2}, Joanna Mattis¹, Aleksandra K. Denisin¹, Sally Pak¹, Hannah Bernstein¹, Charu Ramakrishnan¹, Logan Grosenick¹, Viviana Gradinaru² & Karl Deisseroth^{1,2,3,4}

Obtaining high-resolution information from a complex system, while maintaining the global perspective needed to understand system function, represents a key challenge in biology. Here we address this challenge with a method (termed CLARITY) for the transformation of intact tissue into a nanoporous hydrogel-hybridized form (crosslinked to a three-dimensional network of hydrophilic polymers) that is fully assembled but optically transparent and macromolecule-permeable. Using mouse brains, we show intact-tissue imaging of long-range projections, local circuit wiring, cellular relationships, subcellular structures, protein complexes, nucleic acids and neurotransmitters. CLARITY also enables intact-tissue *in situ* hybridization, immunohistochemistry with multiple rounds of staining and de-staining in non-sectioned tissue, and antibody labelling throughout the intact adult mouse brain. Finally, we show that CLARITY enables fine structural analysis of clinical samples, including non-sectioned human tissue from a neuropsychiatric-disease setting, establishing a path for the transmutation of human tissue into a stable, intact and accessible form suitable for probing structural and molecular underpinnings of physiological function and disease.

The extraction of detailed structural and molecular information from intact biological systems has long been a fundamental challenge across fields of investigation, and has spurred considerable technological innovation^{1–8}. The study of brain structure–function relationships in particular may benefit from intact-system tools^{9–12}, and in general, much valuable information on intra-system relationships and joint statistics will be accessible from full structural analysis of intact systems rather than piecemeal reconstruction across preparations. Yet even tissue structure in itself provides only a certain level of insight without detailed molecular phenotyping^{13,14}, which is difficult to achieve within intact tissue.

Current pioneering methods suitable for the mammalian brain either involve sectioning and reconstruction, or are incompatible with molecular phenotyping, or both. Automated sectioning methods have been successfully used to map structure^{4,5,15–18}, in some cases with molecular labelling. However, detailed reconstruction has typically been limited in application to small volumes of tissue. On the other hand, intact-imaging methods that extend the depth of light microscopy by reducing light scattering have emerged^{19–21}, but these preparations are incompatible with intact-tissue molecular phenotyping, and require many weeks of preparation to achieve partial tissue clearing. Studying intact systems with molecular resolution and global scope remains an unmet goal in biology.

We set ourselves the goal of rapidly transforming intact tissue into an optically transparent and macromolecule-permeable construct while simultaneously preserving native molecular information and structure. We took note of the fact that packed lipid bilayers are implicated in rendering tissue poorly accessible—both to molecular probes and to photons—by creating diffusion-barrier properties relevant to chemical penetration, as well as light-scattering properties at the lipid–aqueous interface^{22,23}. If lipid bilayers could be removed non-destructively, light and macromolecules might penetrate deep into the tissue, allowing three-dimensional imaging and immunohistological

analysis without disassembly. But removing lipid membranes that provide structural integrity and retain biomolecules would inevitably damage tissue with profound loss of cellular and molecular information. Therefore the provision of a physical framework would first be required to physically support the tissue and secure biological information. We have developed and used such a technology, which we term CLARITY, that addresses these challenges.

Hydrogel–electrophoretic tissue transmutation

We began by infusing hydrogel monomers (here, acrylamide and bisacrylamide), formaldehyde and thermally triggered initiators into tissue at 4 °C (Fig. 1). In this step, formaldehyde not only crosslinks the tissue, but also covalently links the hydrogel monomers to biomolecules including proteins, nucleic acids and small molecules. Next, polymerization of the biomolecule-conjugated monomers into a hydrogel mesh was thermally initiated by incubating infused tissue at 37 °C for 3 h, at which point the tissue and hydrogel became a hybrid construct. This hydrogel–tissue hybridization physically supports tissue structure and chemically incorporates biomolecules into the hydrogel mesh. Importantly, lipids and biomolecules lacking functional groups for conjugation remain unbound and therefore can be removed from the hybrid. To extract lipids efficiently, we developed an ionic extraction technique rather than using hydrophobic organic solubilization, for two main reasons. First, although organic solvents can extract lipids or reduce refractive-index variations^{19,20,24}, these solvents quench fluorescence, thereby limiting imaging time. Light-sheet microscopy has been used to image benzyl alcohol/benzyl benzoate (BABB)-treated samples while fluorescence persists^{19,20}, but this approach is less compatible with slower high-resolution imaging. Moreover, instability of native fluorescence in BABB constrains imaging of fine neuronal projections or other modest signals that can be easily quenched. Conversely, in the hydrogel–tissue hybrid all fluorescent proteins tested, including green, yellow and red fluorescent proteins (GFP, YFP and RFP,

¹Department of Bioengineering, Stanford University, Stanford, California 94305, USA. ²CNC Program, Stanford University, Stanford, California 94305, USA. ³Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California 94305, USA. ⁴Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, USA.

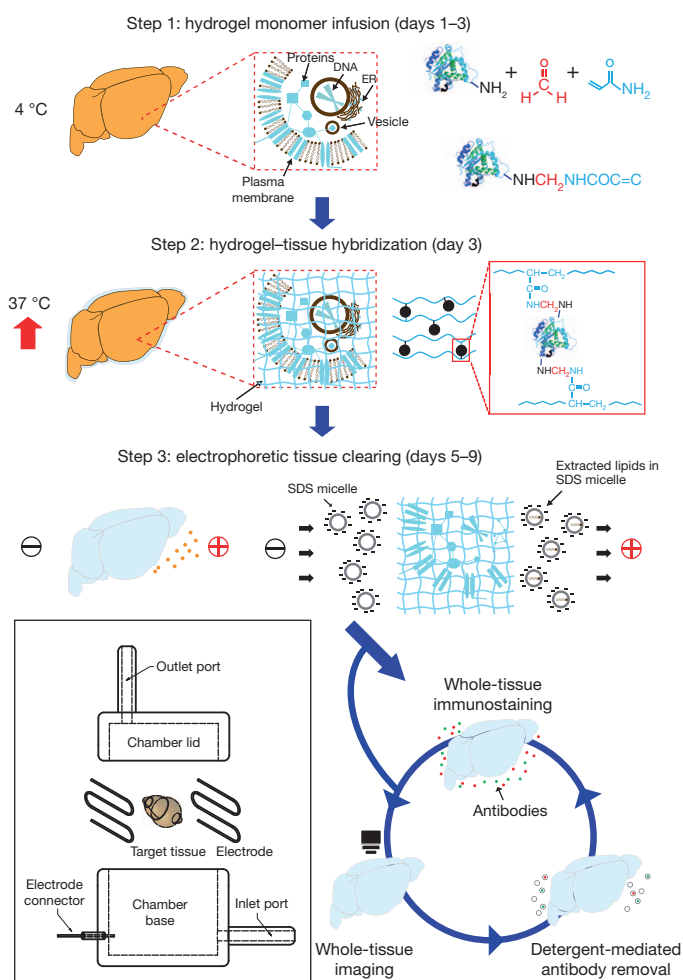


Figure 1 | CLARITY. Tissue is crosslinked with formaldehyde (red) in the presence of hydrogel monomers (blue), covalently linking tissue elements to monomers that are then polymerized into a hydrogel mesh (followed by a day-4 wash step; Methods). Electric fields applied across the sample in ionic detergent actively transport micelles into, and lipids out of, the tissue, leaving fine-structure and crosslinked biomolecules in place. The ETC chamber is depicted in the boxed region (Supplementary Fig. 2).

respectively), were robust to ionic detergent extraction (Fig. 2, Supplementary Fig. 1 and Supplementary Video 1). Second, whereas passive diffusion of detergent micelles would take many months to completely extract lipids from the adult mouse brain, we developed an active-transport organ-electrophoresis approach, which we term electrophoretic tissue clearing (ETC), that capitalizes on the highly charged nature of ionic micelles. This method expedites the extraction by orders of magnitude (Fig. 1 and Supplementary Fig. 2). We anticipated that the hydrogel would secure biomolecules and fine structural features such as membrane-localized proteins, synapses and spines in place, whereas membranous lipids that cause light scattering and prevent penetration of macromolecules would be actively removed, leaving behind an otherwise fully assembled biological system suitable for labelling and imaging intact.

Whole adult mouse brain imaging

To test anticipated features of the technology, we processed with CLARITY (hereafter ‘clarified’) the brain of a 3-month-old line-H mouse, in which the cytosolic fluorescent protein enhanced YFP (eYFP), under control of the Thy1 promoter sequence region, is expressed in a subset of projection neurons^{25,26}. Within 8 days, the intact adult brain was transmuted into a lipid-extracted and structurally stable hydrogel-tissue hybrid. Scattering still occurs owing to

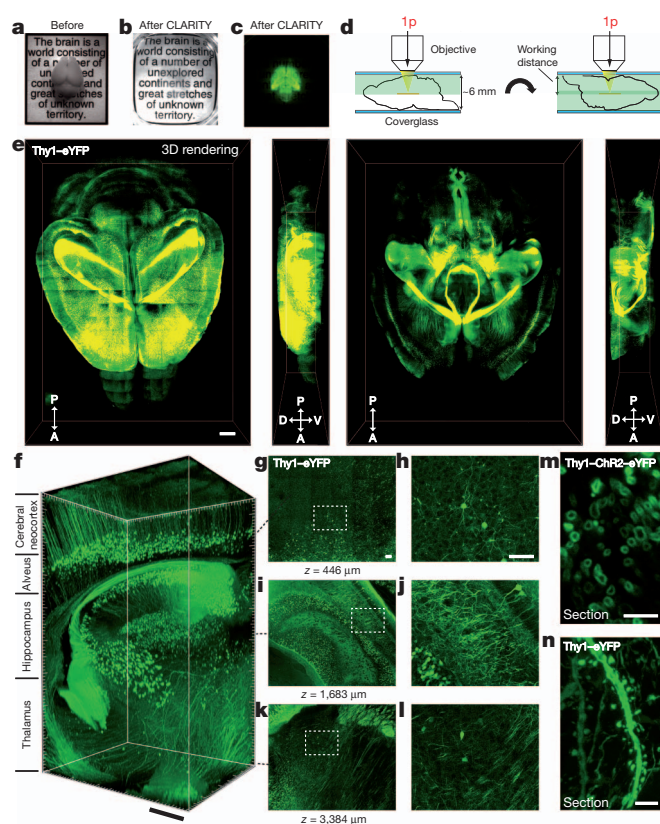


Figure 2 | Intact adult mouse brain imaging. Imaging was performed in adult mouse brains (3 months old). **a**, Cajal quote before CLARITY. **b**, Cajal quote after CLARITY: Thy1-eYFP line-H mouse brain after hydrogel-tissue hybridization, ETC and refractive-index matching (Methods). **c**, Fluorescence image of brain depicted in **b**. **d**, Dorsal aspect is imaged (single-photon (1p) microscopy), then brain is inverted and ventral aspect imaged. **e**, Three-dimensional rendering of clarified brain imaged ($\times 10$ water-immersion objective; numerical aperture, 0.3; working distance, 3.6 mm). Left, dorsal half (stack size, 3,100 μm ; step size, 20 μm). Right, ventral half (stack size, 3,400 μm ; step size, 20 μm). Scale bar, 1 mm (Supplementary Videos 3–5). **f**, Non-sectioned mouse brain tissue showing cortex, hippocampus and thalamus ($\times 10$ objective; stack size, 3,400 μm ; step size, 2 μm). Scale bar, 400 μm (Supplementary Video 2). **g–l**, Optical sections from **f** showing negligible resolution loss even at $\sim 3,400\text{-}\mu\text{m}$ deep: $z = 446\text{ }\mu\text{m}$ (**g**, **h**), $z = 1,683\text{ }\mu\text{m}$ (**i**, **j**) and $z = 3,384\text{ }\mu\text{m}$ (**k**, **l**). **h**, **j** and **l**, boxed regions in **g**, **i** and **k**, respectively. Scale bars, 100 μm . **m**, Cross-section of axons in clarified Thy1-channelrhodopsin2 (ChR2)-eYFP striatum: membrane-localized ChR2-eYFP (1-mm-thick coronal block; $\times 63$ glycerol-immersion objective; numerical aperture, 1.3; working distance, 280 μm). Scale bar, 5 μm . **n**, Dendrites and spines of neurons in clarified Thy1-eYFP line-H cortex (1-mm-thick coronal block; $\times 63$ glycerol objective). Scale bar, 5 μm .

heterogeneously distributed protein and nucleic acid complexes in the hybrid; however, after immersion in refractive-index-specified solutions matching the CLARITY hybrid (for example, 85% glycerol or FocusClear, both with a refractive index of ~ 1.45), the intact brain (including heavily myelinated white matter, thalamus and brainstem) becomes uniformly transparent (Fig. 2a–c and Supplementary Videos 2–5). ETC is crucial: passive refractive-index matching of non-ETC-processed brain provided little clearing (Supplementary Fig. 3), whereas full CLARITY processing enabled imaging of the entire intact mouse brain at cellular resolution even using single-photon microscopy (Fig. 2d–f and Supplementary Videos 2–5); this is important as many dyes and fluorescent proteins are better suited to single-photon rather than two-photon illumination^{27,28}.

Because imaging depth in clarified tissue appeared to be limited only by the working distance of the objective (here 3.6 mm, although 8-mm working distance/0.9 numerical aperture objectives are available),

to image the 5–6-mm-thick adult mouse brain, we next imaged the dorsal half of the brain followed by inversion and imaging of the ventral half (Fig. 2d, e). Figure 2f shows a volume of unsectioned mouse brain with visualization through the cortex, hippocampus and thalamus (Fig. 2g–l and Supplementary Video 2). We observed that tissue expanded with ETC and returned to its original size after refractive-index matching; this transient change did not cause net tissue deformation, and remaining secured in place were fine structural details such as membrane-localized proteins (Fig. 2m), dendritic spines (Fig. 2n) and synaptic puncta (Fig. 3e–g). We found that clarified tissue may be compatible with downstream electron microscopy, retaining key ultrastructural features such as synapses and postsynaptic densities, although with current methods not all relevant ultrastructures are apparent, probably owing to lack of membrane lipids (Supplementary Fig. 4). CLARITY thus allows rapid high-resolution optical access to dense intact tissue and, if needed, subsequent electron microscopy analysis. We next sought to determine whether this design enabled intact-tissue molecular phenotyping.

Molecular phenotyping of intact-tissue volumes

Interrogating molecular features at subcellular resolution in an intact brain with known global wiring properties may be of value; however, conventional labelling techniques involve: (1) loss of native molecules after the permeabilization required for access to intracellular targets²⁹; (2) time-intensive thin sectioning and reconstruction⁵; or (3) damage when multiple rounds of labelling are attempted due to harsh probe-removal processes^{5,30–32}. We anticipated that CLARITY could allow intact-tissue and multiple-round molecular phenotyping by overcoming these three main difficulties. First, CLARITY may preserve native antigens with unusual completeness owing to the hydrogel-hybridization process. To quantify molecular preservation associated with tissue–hydrogel fusion, we compared protein loss in clarified mouse brain to loss from conventional methods (Fig. 3a). Of total protein, ~65% was solubilized (lost) when conventionally paraformaldehyde (PFA)-fixed tissue blocks were cleared by 4% SDS for 1 week. Scale, a tissue-clearing method using 4 M urea, allowed ~41% protein loss over the same interval. Even PFA-fixed tissue treated only with 0.1% Triton X-100 in PBS buffer, a mild detergent-based permeabilization solution used in conventional histology, allowed ~24% loss of protein. However, when hydrogel-hybridized tissue was cleared with the stringent 4% SDS solution of CLARITY, only ~8% protein loss was seen, indicating that chemical tethering of biomolecules into hydrogel mesh can enhance the preservation of molecular tissue components.

Second, we found that CLARITY, which increases tissue permeability by replacing lipid bilayers with nanoporous hydrogel, enabled rapid diffusion of molecular probes deep into intact tissue, and therefore allows access to preserved biomolecules without sectioning. In a 1-mm-thick clarified coronal block of mouse brain, uniformly antibody stained over 3 days (Fig. 3b and Supplementary Video 6), quantitative co-localization analysis revealed that eYFP fluorescence and anti-GFP staining overlapped throughout the block (Fig. 3c, d). We next wanted to know whether whole-mount mouse brain immunolabelling was possible; although the lipid-cleared brain is highly permeable, it was possible that antibodies would not diffuse throughout the tissue on practical timescales. We conducted immunohistochemistry for tyrosine hydroxylase in clarified intact mouse brain, and found that 2 weeks of incubation each for primary and secondary antibody enabled 2.5 mm of penetration from each surface, enough to stain ~5-mm-thick adult mouse brains (Supplementary Fig. 5 and Supplementary Video 7). Tyrosine-hydroxylase-positive cell bodies and projections were readily visualized throughout the brain, and only at the deepest point was moderate signal attenuation seen, indicating that antibody penetration was not quite maximized and may be further enhanced by tuning conditions.

To investigate axonal projections of the tyrosine hydroxylase neurons further, we clarified, tyrosine-hydroxylase-stained and imaged

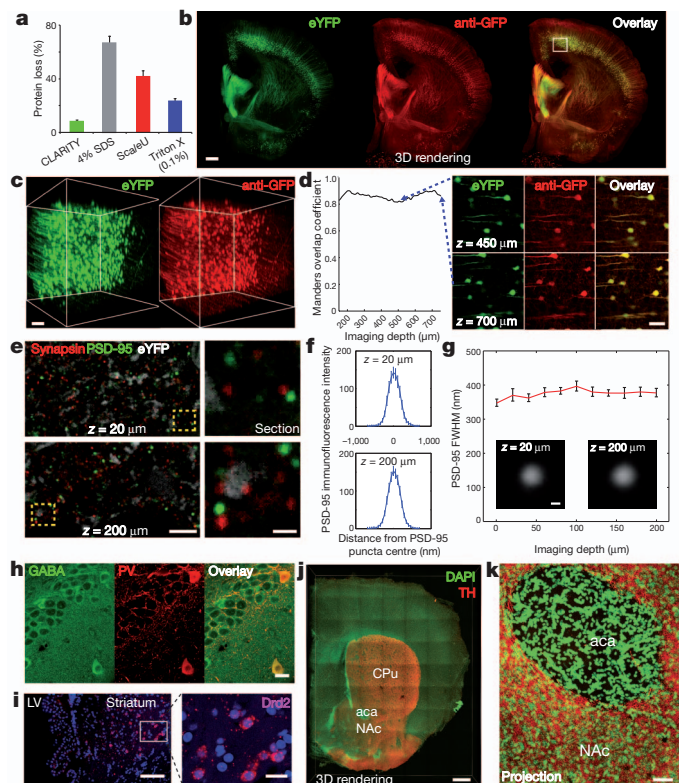


Figure 3 | Molecular phenotyping in intact tissue. **a**, Protein loss in clarified mouse brain compared to conventional methods (see Supplementary Information for more details); error bars denote s.e.m.; $n = 4$ for each condition. **b**, Rendering of a 1-mm-thick non-sectioned coronal block of Thy1-eYFP mouse brain immunostained for GFP. The tissue was ETC-cleared (1 day), immunostained (3 days) and imaged ($\times 10$ water-immersion objective; single-photon excitation). Left, eYFP (green); middle, anti-GFP (red); right, overlay. Scale bar, 500 μm (Supplementary Video 6). **c**, Three-dimensional rendering of the boxed region in **b** shows eYFP fluorescence (left) and anti-GFP staining (right). **d**, Left, co-localization: Manders overlap coefficient plotted versus depth⁴³. Right, optical sections at different depths in three-dimensional rendering. Scale bar, 100 μm . **e**, **f**, 500- μm -thick block of line-H mouse brain (2 months old) clarified for 1 day and immunostained for synapsin I (red) and PSD-95 (blue) for 3 days (Methods) ($\times 63$ glycerol objective; single-photon excitation). **e**, Left, optical sections ($z = 20 \mu\text{m}$, $z = 200 \mu\text{m}$). Right, enlarged images of boxed regions on left. Individual synaptic puncta resolved throughout depth. White depicts eYFP staining. **f**, Average immunofluorescence cross-section of PSD-95 puncta at $z = 20 \mu\text{m}$ (top) and $z = 200 \mu\text{m}$ (bottom). **g**, Full width at half maximum (FWHM) of average immunofluorescence cross-section of PSD-95 puncta versus depth. Insets, average puncta at $z = 20 \mu\text{m}$ and $z = 200 \mu\text{m}$. **h**, Hippocampal staining. Left, GABA; middle, parvalbumin (PV); right, overlay. 500- μm -thick block of wild-type mouse brain (3 months) clarified (1 day) and immunostained (3 days) ($\times 25$ water-immersion objective; numerical aperture, 0.95; working distance, 2.5 mm; single-photon excitation). Scale bar, 20 μm . **i**, *in situ* hybridization. Clarified 500- μm mouse brain block showing dopamine receptor D2 (*Drd2*) mRNA in the striatum. LV, lateral ventricle. Blue, DAPI. 50-base-pair RNA probes for *Drd2* visualized with FastRed ($\times 25$ water-immersion objective; single-photon excitation (555 nm) for FastRed, two-photon excitation (720 nm) for DAPI). Scale bars: left, 100 μm ; right, 20 μm . **j**, **k**, Axonal fibres of tyrosine hydroxylase (TH)-positive neurons in the nucleus accumbens (NAc) and caudate-putamen (CPu). **j**, Three-dimensional rendering of 1-mm-thick clarified mouse brain block stained for tyrosine hydroxylase (red) and DAPI (green). aca, anterior commissure. Scale bar, 500 μm . **k**, Maximum projection, NAc/aca volume in **j**. Scale bar, 50 μm .

1-mm-thick coronal blocks of mouse brain using a high numerical aperture objective (0.95). As shown in Fig. 3j, k, Supplementary Figs 6–8 and Supplementary Video 8, projections of tyrosine-hydroxylase-positive fibres were readily visualized in the neocortex, nucleus

accumbens, caudate-putamen and amygdala. Interestingly, certain cells in the infralimbic cortex and amygdala appeared much more heavily invested by tyrosine hydroxylase fibres than neighbouring cells (Supplementary Figs 6f and 8 and Supplementary Video 8). Further supporting specificity of staining in clarified tissue (Fig. 3j, k and Supplementary Fig. 7), visualized tyrosine hydroxylase projections spared anterior commissure while heavily investing the surrounding nucleus accumbens and caudate-putamen as expected.

We next considered that CLARITY could potentially enable analysis of subcellular molecular architecture in large volumes with resolution at the diffraction limit of light microscopy³³, in an approach complementary to thin mechanical sectioning and three-dimensional reconstruction^{5,34,35}. Resolving submicron-scale structures deep in tissue poses a challenge as the high numerical aperture objectives used for high-resolution imaging are susceptible to optical aberration caused by refractive-index variations in tissue, which blurs images and impairs differentiation of adjacent fluorescent signals, such as pre- and postsynaptic puncta. But when a clarified 500- μ m-thick block of Thy1-eYFP mouse brain was immunolabelled for synapsin I (presynaptic) and PSD-95 (postsynaptic) and imaged with a $\times 63$ glycerol-immersion objective (numerical aperture, 1.3; working distance, 280 μ m), throughout the tissue volume we could unequivocally identify individual paired pre- and postsynaptic puncta, molecularly defining identity and position of putative excitatory synapses (Fig. 3e–g). The full-width at half-maximum of the point-spread function was uniform throughout the block, indicating that loss of resolution is negligible even near the diffraction limit of conventional light microscopy (Fig. 3f, g and Supplementary Figs 9 and 10) and demonstrating another feature of the technique: removal of lipid membranes ensures that tissue refractive index remains nearly constant throughout large volumes, allowing high-resolution imaging. We also found that the CLARITY hydrogel-conjugation process preserves small molecules such as the neurotransmitter GABA (γ -aminobutyric acid, Fig. 3h) and (for *in situ* hybridization) messenger RNAs (Fig. 3i), paving the way for multimodal combinatorial labelling within unsectioned tissue.

Finally, CLARITY was found to enable multi-round molecular phenotyping; the stable framework allowed effective removal of antibodies without fine structural damage or degraded antigenicity. Traditional elution requires incubation in acidic buffer with potassium permanganate (KMnO₄) oxidation or heat, harsh treatments that cause antigen loss and fluorescence quenching^{5,30–32}. But in the hydrogel context of increased structural/chemical stability, the ionic detergent (4% SDS/neutral-pH buffer) used initially to create clarified tissue can be re-used to denature antibodies and disrupt binding, and the permeability of clarified tissue can facilitate rapid antibody exchange. We performed three consecutive rounds of staining in 1-mm-thick coronal blocks from a Thy1-eYFP H-line mouse brain (Fig. 4 and Supplementary Videos 10–14), observing effective antibody removal and preserved eYFP-positive neuronal morphology as well as re-staining capability. Although further quantification would be required to fully map the extent to which CLARITY secures molecular information, this result shows that elution in clarified tissue largely preserves integrity of tissue structure, cellular architecture, fluorescence signals (Fig. 4d–f) and 4',6-diamidino-2-phenylindole (DAPI) DNA staining (Fig. 4g). Moreover, repeated tyrosine hydroxylase staining in the first and third rounds maintained signal pattern and intensity, confirming that antigenicity is retained throughout multiple rounds of staining and elution.

Human tissue imaging and molecular phenotyping

CLARITY is, in principle, applicable to any tissue: adult zebrafish imaging and phenotyping are demonstrated³⁶ in Supplementary Fig. 12, and we also tested post-mortem human brains. With important and rare tissue from clinical brain banks residing in formalin, one concern is that these maximally crosslinked samples could be poorly

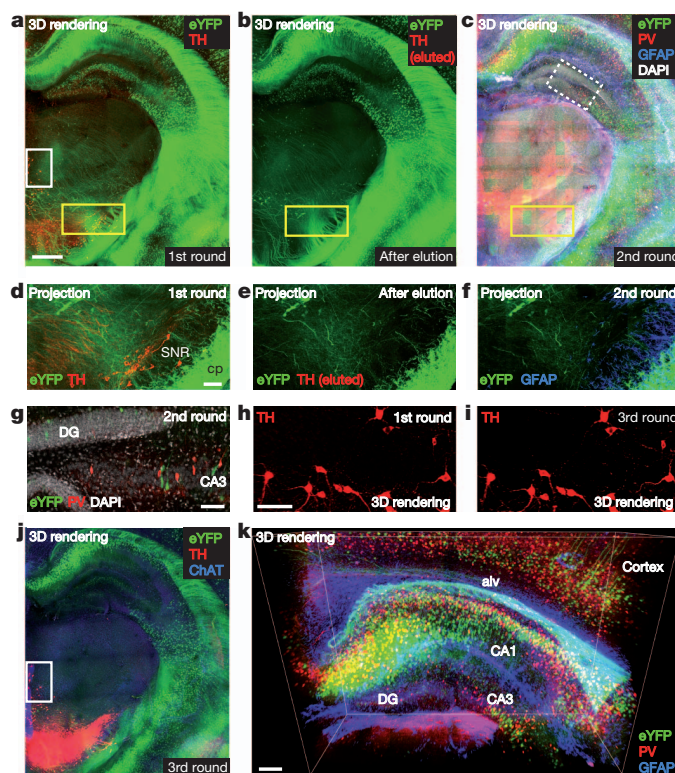


Figure 4 | Multi-round molecular phenotyping of intact tissue. **a**, First round. Rendering of 1-mm-thick Thy1-eYFP block immunostained for tyrosine hydroxylase in non-sectioned form. ETC-cleared (1 day) and immunostained (6 days). Scale bar, 500 μ m (Supplementary Video 10). **b**, Antibodies eluted from block in **a** (4% SDS, 60 °C for 0.5 days). Tyrosine hydroxylase signal was removed and eYFP fluorescence retained (Supplementary Video 11). **c**, Second round. Three-dimensional rendering of same block now immunostained for parvalbumin (red), glial fibrillary acidic protein (GFAP) (blue) and DAPI (white) (Supplementary Video 12). **d–f**, Maximum projections of 100 μ m volume of yellow-boxed regions in **a**, **b** and **c**, respectively. eYFP-positive neurons preserved. cp, cerebral peduncle; SNR, substantia nigra. Scale bar, 100 μ m. **g**, Optical section of white/dotted-box region in **c** showing DAPI. CA, cornu ammonis; DG, dentate gyrus. Scale bar, 100 μ m. **h**, **i**, Tyrosine hydroxylase channel of white box regions in **a** (**h**) and **j** (**i**). Tyrosine hydroxylase antigenicity preserved through multiple elutions. Scale bar, 100 μ m. **j**, Third round. Block in **a–c** immunostained for tyrosine hydroxylase (red) and choline acetyltransferase (ChAT) (blue) (Supplementary Video 13). **k**, Three-dimensional view of hippocampus in **c** showing eYFP-expressing neurons (green), parvalbumin-positive neurons (red) and GFAP (blue). Alv, alveus. Scale bar, 200 μ m (Supplementary Video 14).

suitable for CLARITY, and certainly the mapping of connectivity in the human brain (ref. 37 and <http://www.neuroscienceblueprint.nih.gov/connectome/>) at cellular resolution has been hampered by lack of methods for labelling individual neurites over distances, as the injection of dyes into living brain or *in vivo* genetic labelling are feasible only in animals^{38,39}. If extensively fixed human tissues were suitable for CLARITY, new avenues would be opened for the study of clinical samples.

We found that CLARITY was suitable for such long-banked human brain, allowing immunohistological visualization and identification of neurons and projections over large volumes (Fig. 5a–g and Supplementary Videos 15–17). In 0.5-mm-thick blocks of frontal lobe from an autistic patient, stored in formalin for >6 years, we were able to stain for axons with neurofilament protein and myelin basic protein, and trace individual fibres (Fig. 5e and Supplementary Video 15). In addition, by staining for parvalbumin it was possible to visualize the distribution of parvalbumin-positive neurons in the neocortex over large volumes (6.7 \times 4.7 \times 0.5 mm), and trace individual

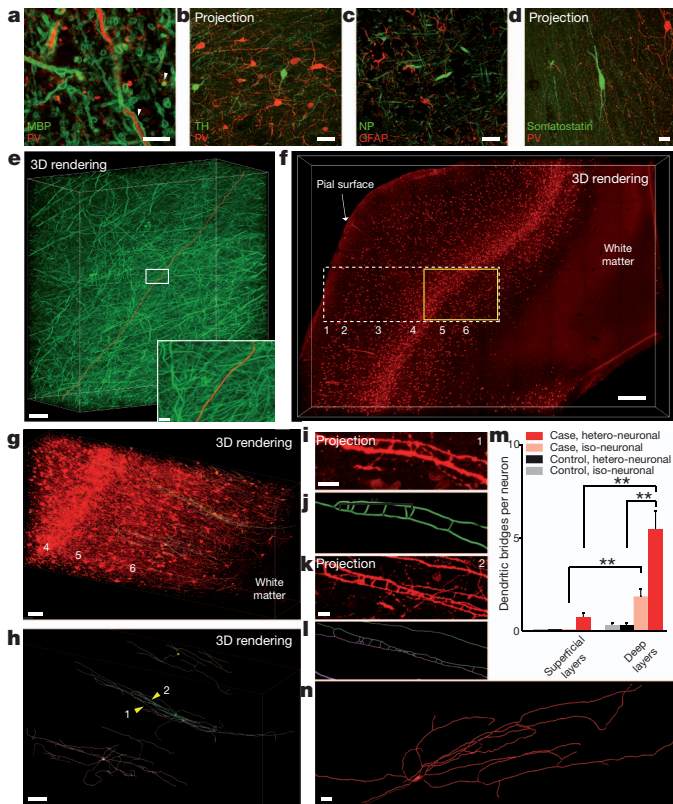


Figure 5 | Human brain structural/molecular phenotyping. Human BA10 500- μm -thick intact blocks clarified (1 day) and immunostained (3 days) ($\times 25$ water-immersion objective). **a**, Optical section: myelin basic protein (MBP) and parvalbumin staining. White arrowheads indicate membrane-localized myelin basic protein around parvalbumin-positive projections. Scale bar, 10 μm . **b**, Tyrosine hydroxylase and parvalbumin staining (maximum projection; 120 μm volume; step size, 0.5 μm). Scale bar, 50 μm . **c**, Optical section: neurofilament (NP) and GFAP. Scale bar, 20 μm . **d**, Somatostatin and parvalbumin staining (maximum projection; 63 μm volume; step size, 0.5 μm). Scale bar, 20 μm . **e**, Rendering of neurofilament-positive axonal fibres. Red, traced axon across volume. Scale bar, 500 μm . Inset: boxed region. Scale bar, 20 μm (Supplementary Video 15). **f**, Visualization of parvalbumin-positive neurons in the neocortex of autism case; layers identified as described in ref. 44. Scale bar, 500 μm (Supplementary Video 16). **g**, Yellow-boxed region in **f** showing parvalbumin-positive cell bodies and fibres in layers 4, 5 and 6. Three representative parvalbumin-positive interneurons in layer 6 with ladder-shaped hetero- or iso-neuronal connections were traced (green, purple, blue). Scale bar, 100 μm (Supplementary Video 17). **h**, Three-dimensional rendering of abnormal neurons in **g**; yellow arrowheads (1, 2) indicate ladder-shaped structures shown below in **i** and **k**. Scale bar, 80 μm . **i**, Zoomed-in maximum projection of 8 μm volume showing morphology of ladder-shaped structure formed by neurites from a single neuron. Scale bar, 10 μm . **j**, Tracing of structure in **i**. **k**, Maximum projection of 18 μm volume showing ladder-shaped structure formed by neurites from two different neurons. Scale bar, 10 μm . **l**, Tracing of structure in **k**. **m**, Iso- and hetero-neuronal dendritic bridges per neuron. Neurons selected randomly and traced in software (Methods); dendritic bridges were manually counted. $**P < 0.05$; error bars denote s.e.m. $n = 6$ neurons for both superficial and deep layers of autism case and $n = 4$ neurons for both superficial and deep layers of control case. **n**, Three-dimensional reconstruction of a neuron in layer 2 (superficial) of the autism case. Typical avoidance of iso-dendritic contact was observed.

parvalbumin-labelled processes (Fig. 5g–n and Supplementary Videos 16 and 17).

Unlike mechanical sectioning methods that may involve deformation of tissue and uncertainty in registration across sections, CLARITY preserves continuity of structure, which not only allows tracing of neurites over distances, but also provides a class of distinct

information about three-dimensional and topological morphology of traced neurons. As one example, we found that many parvalbumin-positive interneurons in this human sample, particularly in deep layers, showed isoneuronal and heteroneuronal dendritic bridges (Fig. 5g–m, Supplementary Fig. 13, Supplementary Video 17 and Supplementary Table 2). These ladder-like connections, not typical of age- and sex-matched normal brain (Supplementary Fig. 14), instead resemble abnormalities observed with mutations in the Down syndrome cell-adhesion molecule (Dscam) protein⁴⁰ or protocadherins⁴¹; mutations in the latter family are associated with autism-spectrum disorder⁴². Although extensive work would be required to define the incidence and implications of abnormal parvalbumin neurons, the observation illustrates the structural and molecular capability that clarified tissue provides by virtue of volumetric continuity, which may help to shed light on structural underpinnings of neuropsychiatric disease.

Discussion

With this hydrogel–tissue fusion and electrophoretic-clearing technology, intact tissue can be rapidly transformed into an optically and chemically accessible form while retaining structural and molecular information, thereby enabling the imaging of entire adult vertebrate brains as well as multiple-round molecular phenotyping without thin sectioning. Currently, the non-corrected water-immersion objectives used, rather than clarified tissue properties, represent the main limiting factor for maximizing imaging depth and resolution; specialized adaptive optics or CLARITY-optimized long-working-distance objectives will further enhance the acquisition of integrated structural and molecular information from intact systems.

Turning immense data sets into useful insights remains a key challenge. Computational approaches to image segmentation, three-dimensional registration and automated tracing require further development. Moreover, for the subset of applications relevant to circuit connectivity, CLARITY does not in itself supplant electron microscopy. However, the potential electron microscopy compatibility of CLARITY (Supplementary Fig. 4) may offer a two-step approach in probing connectivity with synaptic resolution; CLARITY with light microscopy may first be used to visualize and trace fluorescently labelled neurons and projections in the whole brain, followed by electron microscopy analysis of small volumes to define the patterns and rules (for example, postsynaptic cell target type) followed by axon terminals and synaptic contacts. Notably, owing to the absence of lipids, osmium tetroxide/uranyl acetate staining alone with electron microscopy does not provide enough contrast to identify all ultrastructural boundaries, and additional methodological tissue-preparation work is indicated.

Although much remains to be refined, newly enabled access to intact-system properties may dovetail with existing optogenetic-control or activity-imaging methods. For example, cells studied in local and global structural detail with CLARITY could be the same cells previously visualized for activity, or controlled to modulate performance, during behavioural assays in the same organism with proper registration. Moreover, efficient molecular phenotyping applied to immediate early gene products may help to map populations with altered activity downstream of directly modulated populations, across the same intact brain and in the context of global wiring maps. Together with its capability for intersectional definition through multiple rounds of labelling, and broad application domain (including mouse, zebrafish and human), these data suggest that CLARITY provides access to structural and molecular information that may help to support integrative understanding of large-scale intact biological systems.

METHODS SUMMARY

CLARITY. CLARITY refers to transformation of intact biological tissue into a hybrid form in which specific components are replaced with exogenous elements that provide new accessibility or functionality. Initially, the term was an acronym to describe the Clear Lipid-exchanged Acrylamide-hybridized Rigid Imaging/Immunostaining/*In situ* hybridization-compatible Tissue-hydrogel. However, CLARITY is more inclusively defined because infused elements need not be

exclusively hydrogel monomers or acrylamide-based, and the properties of infused elements may be adjusted for varying degrees of clarity, rigidity, macromolecule-permeability or other functionality.

CLARITY for mouse brain. Adult mice (4–12 weeks old) were anaesthetized with Beuthenasia-D and perfused transcardially with a mixture of 4% (wt) PFA, 4% (wt/vol) acrylamide, 0.05% (wt/vol) bis-acrylamide, 0.25% (wt/vol) VA044 and PBS. Brains were then extracted and incubated in the same solution at 4 °C for 3 days. Solution temperature was then increased to 37 °C to initiate polymerization. After 3 h at 37 °C, hydrogel-embedded brains were placed in a custom-built organ-electrophoresis chamber. While sodium borate buffer (200 mM, pH 8.5) containing 4% (wt/vol) SDS was circulated through the chamber, 10–60 V was applied across the brains at 37–50 °C for 2 days. After clearing, brains were incubated in PBS at 37 °C for 2 days to remove SDS. To prepare 1-mm-thick coronal blocks of mouse brain for immunostaining, hydrogel-embedded and clarified brains were cut into 1-mm-thick blocks using a mouse brain matrix (Ted Pella). Blocks were then cleared by electrophoresis for 1 day as described above.

CLARITY for human brain. Frontal lobe (BA10) from post-mortem human brain tissue was sliced into 500-µm-thick blocks using vibratome. Blocks were incubated in 4% (wt) PFA, 4% (wt/vol) acrylamide, 0.25% (wt/vol) VA044 and PBS at 4 °C for 1 week. The temperature of the solution was then increased to 37 °C to initiate polymerization. After 3 h incubation at 37 °C, hydrogel-embedded tissues were placed in the custom-built electrophoresis chamber. While sodium borate buffer (200 mM, pH 8.5) containing 4% (wt/vol) SDS circulated through the chamber, 10–60 V was applied at 37–50 °C for 1 day. Tissues were then incubated in PBS at 37 °C for 1 day to remove SDS.

Full Methods and any associated references are available in the online version of the paper.

Received 1 September 2012; accepted 20 March 2013.

Published online 10 April 2013.

- Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73–76 (1990).
- Carmeliet, P. & Tessier-Lavigne, M. Common mechanisms of nerve and blood vessel wiring. *Nature* **436**, 193–200 (2005).
- Helmchen, F. & Denk, W. Deep tissue two-photon microscopy. *Nature Methods* **2**, 932–940 (2005).
- Denk, W. & Horstmann, H. Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol.* **2**, e329 (2004).
- Micheva, K. D. & Smith, S. J. Array tomography: a new tool for imaging the molecular architecture and ultrastructure of neural circuits. *Neuron* **55**, 25–36 (2007).
- Livet, J. *et al.* Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007).
- Li, A. *et al.* Micro-optical sectioning tomography to obtain a high-resolution atlas of the mouse brain. *Science* **330**, 1404–1408 (2010).
- Botcherby, E. J. *et al.* Aberration-free three-dimensional multiphoton imaging of neuronal activity at kHz rates. *Proc. Natl Acad. Sci. USA* **109**, 2919–2924 (2012).
- DeFelipe, J. From the connectome to the synaptome: an epic love story. *Science* **330**, 1198–1201 (2010).
- Petersen, C. C. H. The functional organization of the barrel cortex. *Neuron* **56**, 339–355 (2007).
- Kasthuri, N. & Lichtman, J. W. The rise of the “projectome”. *Nature Methods* **4**, 307–308 (2007).
- Lichtman, J. W., Livet, J. & Sanes, J. R. A technicolour approach to the connectome. *Nature Rev. Neurosci.* **9**, 417–422 (2008).
- Mombaerts, P. *et al.* Visualizing an olfactory sensory map. *Cell* **87**, 675–686 (1996).
- Insel, T. R. & Young, L. J. The neurobiology of attachment. *Nature Rev. Neurosci.* **2**, 129–136 (2001).
- Micheva, K. D., Busse, B., Weiler, N. C., O'Rourke, N. & Smith, S. J. Single-synapse analysis of a diverse synapse population: proteomic imaging methods and markers. *Neuron* **68**, 639–653 (2010).
- Bock, D. D. *et al.* Network anatomy and *in vivo* physiology of visual cortical neurons. *Nature* **471**, 177–182 (2011).
- Briggman, K. L., Helmstaedter, M. & Denk, W. Wiring specificity in the direction-selectivity circuit of the retina. *Nature* **471**, 183–188 (2011).
- Ragan, T. *et al.* Serial two-photon tomography for automated *ex vivo* mouse brain imaging. *Nature Methods* **9**, 255–258 (2012).
- Dodd, H. U., Leischner, U. & Schierloh, A. Ultramicroscopy: three-dimensional visualization of neuronal networks in the whole mouse brain. *Nature Methods* **4**, 331–336 (2007).
- Ertürk, A., Mauch, C., Hellal, F. & Förstner, F. Three-dimensional imaging of the unsectioned adult spinal cord to assess axon regeneration and glial responses after injury. *Nature Med.* **18**, 166–171 (2012).
- Hama, H. *et al.* Scale: a chemical approach for fluorescence imaging and reconstruction of transparent mouse brain. *Nature Neurosci.* **14**, 1481–1488 (2011).
- Cheong, W., Prahl, S. & Welch, A. A review of the optical properties of biological tissues. *IEEE J. Quant. Electron.* **26**, 2166–2185 (1990).
- Sykova, E. & Nicholson, C. Diffusion in brain extracellular space. *Physiol. Rev.* **88**, 1277–1340 (2008).
- Becker, K., Jähring, N., Saghati, S., Weiler, R. & Dodd, H.-U. Chemical clearing and dehydration of GFP expressing mouse brains. *PLoS ONE* **7**, e33916 (2012).
- Porrero, C., Rubio-Garrido, P., Avendaño, C. & Clascá, F. Mapping of fluorescent protein-expressing neurons and axon pathways in adult and developing Thyl1-eYFP-H transgenic mice. *Brain Res.* **1345**, 59–72 (2010).
- Feng, G. *et al.* Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *Neuron* **28**, 41–51 (2000).
- Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing fluorescent proteins. *Nature Methods* **2**, 905–909 (2005).
- Drobizhev, M., Makarov, N. S., Tillo, S. E., Hughes, T. E. & Rebane, A. Two-photon absorption properties of fluorescent proteins. *Nature Methods* **8**, 393–399 (2011).
- Goldenthal, K. L., Hedman, K., Chen, J. W., August, J. T. & Willingham, M. C. Postfixation detergent treatment for immunofluorescence suppresses localization of some integral membrane proteins. *J. Histochem. Cytochem.* **33**, 813–820 (1985).
- Tramu, G. An efficient method of antibody elution for the successive or simultaneous localization of two antigens by immunocytochemistry. *J. Histochem. Cytochem.* **26**, 322–324 (1978).
- Wahlby, C., Erlandsson, F., Begntsson, E. & Zetterberg, A. Sequential immunofluorescence staining and image analysis for detection of large numbers of antigens in individual cell nuclei. *Cytometry* **71**, 32–41 (2002).
- Kolodziejczyk, E. & Baertschi, A. J. Multiple immunolabeling in histology: a new method using thermo-inactivation of immunoglobulins. *J. Histochem. Cytochem.* **34**, 1725–1729 (1986).
- Zhou, Q., Homma, K. J. & Poo, M. Shrinkage of dendritic spines associated with long-term depression of hippocampal synapses. *Neuron* **44**, 749–757 (2004).
- Harlow, M. L., Rens, D., Stoschek, A., Marshall, R. M. & McMahan, U. J. The architecture of active zone material at the frog's neuromuscular junction. *Nature* **409**, 479–484 (2001).
- Chen, X. *et al.* Organization of the core structure of the postsynaptic density. *Proc. Natl Acad. Sci. USA* **105**, 4453–4458 (2008).
- Macdonald, R. Zebrafish immunohistochemistry. *Methods Mol. Biol.* **127**, 77–88 (1999).
- Sporns, O., Tononi, G. & Kötter, R. The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* **1**, e42 (2005).
- Wickersham, I. R., Finke, S., Conzelmann, K.-K. & Callaway, E. M. Retrograde neuronal tracing with a deletion-mutant rabies virus. *Nature Methods* **4**, 47–49 (2007).
- Miyamichi, K. *et al.* Cortical representations of olfactory input by trans-synaptic tracing. *Nature* **472**, 191–196 (2011).
- Matthews, B. J. *et al.* Dendrite self-avoidance is controlled by *Dscam*. *Cell* **129**, 593–604 (2007).
- Lefebvre, J. L., Kostadinov, D., Chen, W. V., Maniatis, T. & Sanes, J. R. Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. *Nature* **488**, 517–521 (2012).
- Morrow, E. M. *et al.* Identifying autism loci and genes by tracing recent shared ancestry. *Science* **321**, 218–223 (2008).
- Dunn, K. W., Kamocka, M. M. & McDonald, J. H. A practical guide to evaluating colocalization in biological microscopy. *Am. J. Physiol. Cell Physiol.* **300**, C723–C742 (2011).
- Condé, F., Lund, J. S., Jacobowitz, D. M., Baimbridge, K. G. & Lewis, D. A. Local circuit neurons immunoreactive for calretinin, calbindin D-28k or parvalbumin in monkey prefrontal cortex: distribution and morphology. *J. Comp. Neurol.* **341**, 95–116 (1994).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was funded by a National Institutes of Health (NIH) Director's Transformative Research Award (TR01) to K.D. from NIMH, as well as by NSF, the Simons Foundation, and the President and Provost of Stanford University. K.D. is also funded by NIDA, the DARPA REPAIR program, and the Wieggers, Snyder, Reeves, Gatsby, and Yu Foundations. K.C. is supported by the Burroughs Wellcome Fund Career Award at the Scientific Interface. S.-Y.K. is supported by a Samsung Scholarship, A.S.A. by the Helen Hay Whitney Foundation, K.A.Z. and A.K.D. by an NSF Graduate Research Fellowship and J.M. by the NIH MSTP. We acknowledge H. Vogel, L. Luo, L. Schwarz, M. Monje, S. Hestrin and D. Castaneda for advice, and the Autism Tissue Program for providing human brain tissue, as well as J. J. Perrino, J. Mulholland and the Cell Sciences Imaging Facility at Stanford for electron microscopy imaging and advice. We would also like to thank the entire Deisseroth laboratory for discussions and support. CLARITY resources and protocols are freely supported online (<http://CLARITYresourcecenter.org>).

Author Contributions K.C. and K.D. conceived and designed the experiments and wrote the paper. K.C. led development of the CLARITY technology and its implementation. K.C. and S.K. clarified samples. K.C. imaged samples. K.C., S.-Y.K., S.K., J.W., K.A.Z., S.P., J.J.M., J.M., V.G. and H.B. prepared animals. J.W. performed tracing. T.J.D. and A.S.A. performed image processing. K.C., J.W., J.J.M. and A.K.D. wrote the CLARITY protocol. A.K.D. created AutoCAD drawings. K.C. and L.G. created Supplementary Videos. A.S.A. contributed to *in situ* and zebrafish data. C.R., L.G. and V.G. contributed to set-up of the relevant laboratory infrastructure. K.D. supervised all aspects of the work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.D. (deisseroth@stanford.edu).

METHODS

CLARITY. CLARITY refers to transformation of intact biological tissue into a hybrid form in which specific components are replaced with exogenous elements that provide new accessibility or functionality. Initially, the term was an acronym to describe the Clear Lipid-exchanged Acrylamide-hybridized Rigid Imaging/Immunostaining/*In situ* hybridization-compatible Tissue-hYdrogel. However, CLARITY is more inclusively defined because infused elements need not be exclusively hydrogel monomers or acrylamide-based, and the properties of infused elements may be adjusted for varying degrees of clarity, rigidity, macromolecule-permeability or other functionality.

Immunostaining of CLARITY-processed mouse brain tissue. For GFP staining shown in Fig. 3b–d, the clarified 1-mm-thick block of Thy1-eYFP line-H mouse (2 months old) brain was incubated at 37 °C for 2 days in 0.1% Triton X-100 (wt/vol), anti-GFP antibody conjugated with Alexa Fluor 594 (Invitrogen, 1:50 dilution), 1 M sodium borate buffer solution, pH 8.5, followed by wash at 37 °C for 1 day with 0.1% (wt/vol) Triton X-100 and 1 M sodium borate buffer, pH 8.5. For tyrosine hydroxylase staining of whole mouse brain in Supplementary Fig. 4, the clarified C57BL/6 mouse (4 weeks old) brain was incubated at 37 °C for 2 weeks in 0.1% (wt/vol) Triton X-100, primary antibody (dilution, 1:50), 0.5 M sodium borate buffer solution, pH 8.5, followed by wash at 37 °C for 1 week with 0.1% (wt/vol) Triton X-100 and 0.5 M sodium borate buffer, pH 8.5. The tissue was then incubated at 37 °C for 2 weeks in 0.1% (wt/vol) Triton X-100, secondary antibody (dilution, 1:50) and 0.5 M sodium borate buffer solution, pH 8.5, followed by wash at 37 °C for 1 week with 0.1% (wt/vol) Triton X-100 and 0.5 M sodium borate buffer, pH 8.5. For other staining, the clarified 1-mm-thick block of Thy1-eYFP H-line mouse (2 months old) brain was incubated at 37 °C for 2 days in 0.1% (wt/vol) Triton X-100, primary antibody (dilution, 1:50–1:100) and 1 M sodium borate buffer solution, pH 8.5, followed by wash at 37 °C for 1 day with 0.1% (wt/vol) Triton X-100 and 1 M sodium borate buffer, pH 8.5. The tissue was then incubated at 37 °C for 1 day in 0.1% (wt/vol) Triton X-100, secondary antibody (dilution, 1:50–1:100) and 1 M sodium borate buffer solution, pH 8.5, followed by wash at 37 °C for 1 day with 0.1% (wt/vol) Triton X-100 and 1 M sodium borate buffer, pH 8.5.

Immunostaining of CLARITY-processed post-mortem human brain tissue. Clarified 500-µm-thick blocks were incubated at 37 °C for 1 day in 0.1% (wt/vol) Triton X-100, primary antibody (dilution, 1:50–1:100) and 1 M sodium borate buffer solution, pH 8.5, followed by wash at 37 °C for 12 h with 0.1% (wt/vol) Triton X-100 and 1 M sodium borate buffer, pH 8.5. The tissue was then incubated at 37 °C for 1 day in 0.1% (wt/vol) Triton X-100, secondary antibody (dilution, 1:50–1:100) and 1 M sodium borate buffer solution, pH 8.5, followed by wash at 37 °C for 12 h with 0.1% (wt/vol) Triton X-100 and 1 M sodium borate buffer, pH 8.5.

Imaging of CLARITY-processed mouse brain tissue. For imaging the intact Thy1-eYFP H-line mouse brain (Fig. 2e), the clarified brain (3 months old) was incubated in FocusClear, a water-based immersion medium, for 2 days. The brain was then enclosed between two coverglass-bottom Petri dishes. The dorsal half of the brain was first imaged (stack size, 3.1 mm; step size, 20 µm) using a Leica SP5 system equipped with the ×10 water-immersion objective (Leica HCX APO L; numerical aperture, 0.30; working distance, 3.6) and 514-nm excitation. The mounted brain was then inverted and the ventral half was imaged in the same way. To obtain the 3.4-mm-thick volume image visualizing from the cortex to the thalamus (Fig. 2f–l), the intact line-H mouse brain was mounted as described and imaged using the ×10 objective and 514-nm excitation (stack size, 3.4 mm; step size, 2 µm). The immunostained 1-mm-thick coronal blocks were incubated in FocusClear for 1 day and enclosed between the coverglass-bottom Petri dish and slide glass. The mounted samples were imaged using the ×10 objective and single-photon excitation (514 nm, 591 nm and 654 nm).

Imaging of CLARITY-processed post-mortem human brain tissue. Clarified and immunostained tissues were incubated in FocusClear for 1 day and mounted as described above. The tissues were then imaged using the Leica SP5 system equipped with the ×25 water-immersion objective (Leica HCX IRAPO L; numerical aperture, 0.95; working distance, 2.4 mm; stack size, 500 µm; step size, 0.5 µm).

Protein-loss measurement. Six PFA-fixed adult mouse (4 weeks old) brains (for 4% SDS, Scale and Triton X-100 treatments) and two (4 weeks old) PFA-fixed, hydrogel-embedded brains (for CLARITY) were cut into 1-mm-thick coronal blocks. One half of each PFA-fixed brain was weighed and placed in 2.5 ml of 4% SDS, ScaleU2 (a mixture of 4 M urea and 30% glycerol) or 0.1% Triton X-100 solution. One half of each PFA-fixed, hydrogel-embedded brain was placed in 2.5 ml of 4% SDS solution. The tissues were allowed to clear for 1 week at 37 °C in the respective solutions and quantity of protein lost from tissue by diffusing into solution was measured using the BCA (bicinchoninic acid) protein assay; total protein in mouse brain was estimated at 10% (wt) on the basis of previous work^{45,46}.

Experimental subjects. Animal husbandry and all aspects of animal care and euthanization as described were in accordance with guidelines from the National Institutes of Health and have been approved by members of the Stanford Institutional Animal Care and Use Committee. Use of post-mortem human tissue was in accordance with guidelines from the National Institutes of Health and approved by the Stanford Institutional Review Board.

Neurite tracing. Manual tracing of individual neurons was performed using Imaris software (Bitplane). Neurons with cell bodies located in the middle 150 µm of the z-stack were randomly sampled and chosen for tracing. The cell body was reconstructed semi-automatically through a user-defined threshold that included as much of the cell body as possible but less than 5 µm of any dendrite. All neurites originating from the cell body were traced manually in short segments in the 'Surpass' mode, and each segment was automatically centred (with the opportunity for user corrections) before being connected. The number of interconnections between filaments of the same cell, and the number of interconnections between a filament from the traced cell and that of any other cell, was counted manually. Separately, for Fig. 5h–n neurons with interesting structures were chosen non-randomly and traced using the same method described above.

Basic CLARITY protocol. Hydrogel solution preparation: (1) Combine and mix 40 ml of acrylamide (40%), 10 ml of bis-acrylamide (2%), 1 g of VA-044 initiator (10% wt), 40 ml of ×10 PBS, 100 ml of 16% PFA and 210 ml of dH₂O with special attention to temperature and safety precautions. Keep all components on ice at all times to prevent polymerization. Caution: PFA is an irritant, sensitizer, carcinogen and toxin. Acrylamide is a potent neurotoxin, a respiratory and skin sensitizer, carcinogen, irritant, mutagen, teratogen and reproductive hazard. Many of the chemical constituents of hydrogels that could be used for CLARITY would fall into one or more of these categories. Therefore, to avoid skin contact or inhalation of monomers and/or crosslinkers (for example, acrylamide or PFA), solution preparation and all subsequent handling of hydrogel solution and polymer must be conducted in a fume hood with personal protective equipment including face shield, laboratory coat, gloves and closed-toe shoes.

Saponin is a widely used mild non-ionic surfactant often used to permeabilize cellular membranes in conventional immunohistochemistry. In CLARITY, saponin can be used in the hydrogel monomer infusion process to facilitate diffusion of the hydrogel monomer and initiator into the tissue, particularly for samples in which cardiac perfusion is not feasible, such as post-mortem human tissues and zebrafish brains. Saponin shortens incubation time required in the hydrogel monomer infusion process. However, bubbles may form that could be linked to saponin use, so routine saponin is not suggested. (2) Distribute 40-ml aliquots into 50-ml conical tubes on ice. Each tissue sample will require the use of one 40-ml tube: 20 ml for perfusion and the remaining 20 ml for sample embedding. (3) Seal tubes tightly and keep in secondary containment (on ice) before removing them from the hood. Transfer aliquots from ice to –20 °C. Store these solutions at –20 °C until they are ready to be used. They can be stored at –20 °C indefinitely if all components were kept sufficiently cold during the preparation process.

Solution preparation: (1) Combine and mix 123.66 g boric acid, 400 g sodium dodecyl sulphate, and 9 l dH₂O. To avoid skin contact or inhalation, prepare solution in a fume hood in proper PPE. Paying special attention to safety, combine water, boric acid and SDS while stirring. Add dH₂O to 10 l and add NaOH until the pH has reached 8.5. This solution can be made, stored and used at room temperature (20 °C). Caution: SDS is a toxin and irritant to the skin and respiratory system.

Transcardial perfusion with hydrogel solution: (1) Before perfusing, thaw the frozen hydrogel monomer solution in the refrigerator or on ice. (2) When the solution is completely thawed and transparent (but still ice cold), gently invert to mix. Make sure that there is no precipitate and avoid introducing any bubbles into the solution. (3) Prepare perfusion materials within a fume hood. (4) Deeply anaesthetize adult mouse with Beuthanasia-D. (5) Perfuse the animal transcardially, first with 20 ml of ice-cold 1× PBS and then 20 ml of the ice-cold hydrogel solution. Maintain a slow rate of perfusion (about 2 min for the 20 ml of each solution). (6) Immediately place the tissue (for example, brain) in 20 ml cold hydrogel solution in a 50-ml conical tube. Keep the sample in solution on ice until it can be moved to a 4 °C refrigerator. (7) Cover sample in aluminium foil if it contains fluorophores and incubate at 4 °C for 2–3 days to allow for further diffusion of the hydrogel solution into the tissue.

Hydrogel tissue embedding: (1) De-gas the 50-ml conical tube containing the sample in the desiccation chamber (in a fume hood) to replace all of the gas in the tube with nitrogen (as oxygen impedes hydrogel formation), as follows: place the 50-ml conical tube on a rack in the desiccation chamber; twist the 50-ml conical tube open sufficiently to allow gas exchange; turn on the nitrogen tank and adjust the control valve such that the inlet to the desiccation chamber fills with nitrogen

gas; switch the desiccation chamber valve from nitrogen gas flow to the vacuum. Turn on the vacuum pump; verify that the chamber is under full vacuum by testing the chamber lid. Keep vacuum on for 10 min; switch the vacuum off and slowly turn the valve to fill the chamber with nitrogen gas; carefully open the chamber just enough to reach the tubes while purging with nitrogen gas, taking great care to minimize exposure to air, and quickly and tightly close the sample tube. (2) Submerge the tube in 37 °C water bath in a 37 °C room or incubator on the rotator. Incubate for 3 h or until solution has polymerized. (3) In a fume hood, extract the embedded sample from the gel (carefully take the sample out and remove extra gel pieces with gloved fingers). Hydrogel waste disposal should be conducted in accordance with all institutional, state and country regulations for hydrogel monomers and crosslinkers (for example, acrylamide and PFA). (4) Wash the sample with 50 ml clearing solution for 24 h at room temperature to dialyse out extra PFA, initiator and monomer. Wash the sample twice more with 50 ml for 24 h, each at 37 °C to further reduce residual PFA, initiator and monomer. Take care to dispose of this waste solution carefully as a biohazard.

ETC: (1) Construct the ETC chamber as described (<http://CLARITYresource.center.org>): place the sample in the chamber, and circulate the clearing solution through the chamber using the temperature-controlled water circulator, with 10–60 V applied across the tissue (for example, brain) at 37–50 °C for several days to clear the sample. The clearing process is faster at higher voltage and temperature, but requires more power, limiting the number of clearing set-ups simultaneously operable by one power supply (typical power output maximum, 300 W). For example, four set-ups can be run simultaneously at 37 °C and 30 V, whereas only two set-ups can be run at 50 °C and 60 V, so circulator temperature and voltage should be chosen to meet practical considerations. In addition, too-high voltage operation could cause bubble formation in the tissue and deposit of black particles on the surface of tissue. Therefore, low voltage (10–40 V) is recommended. Note that lipids and biomolecules lacking functional groups required for conjugation, such as native phosphatidylinositol 4,5-bisphosphate or exogenous dextrans used for labelling, may be lost during this process. (2) After several days, wash the sample twice with of PBST (0.1% Triton X-100 in 1× PBS) twice for 24 h each.

Preparing the sample for imaging: (1) Incubate the sample in FocusClear or 80% glycerol solution for 2 days before imaging; these have refractive indices matching that of clarified tissue. Ensure there is sufficient solution surrounding the sample, and that evaporative losses do not occur. Protect the sample from light. (2) Take a clean glass slide and gently place it on a dust-free surface. (3) Take a small piece of BluTack putty and prepare constant-diameter worm shapes using gloved hands. Make the thickness uniform and about 1.5× the thickness of the sample (for example, if the sample is 1 mm, make the putty tube diameter 1.5 mm). (4) Place two tubes of putty horizontally across the vertical slide, leaving space in between for the tissue sample. Cut excess putty that protrudes off the slide. (5) Using a spatula, carefully take the sample and place it between the putty tubes in the middle of the slide. Pipette ~20 µl of FocusClear medium on top of the sample. (6) Carefully place a Willco dish (with the lipped side facing up) on top of the putty tubes. Press down on the glass part of the dish (keeping fingers over the putty to avoid glass shattering) carefully and slowly until contact is made with the sample and FocusClear medium. Ensure that there are no bubbles between the sample, medium, slide and dish. (7) Using a P200 pipette, carefully introduce more FocusClear to either side of the sealed chamber (from the liquid that surrounded the sample for incubation as it has been optically matched). Take great care not to introduce any bubbles. (8) Now that the whole chamber is filled with FocusClear, use PDMS sealant (Kwik-Sil) across both vertical openings between the putty, dish and slide to fully seal the chamber and prevent evaporation. (9) Place aluminium foil on top of the chamber and place it on a level surface (shielded from light to minimize photodamage). Leave the sample for 10–15 min to allow the PDMS sealant to polymerize fully. (10) Preparation is now ready for imaging.

45. Cheng, C. M. *et al.* Biochemical and morphometric analyses show that myelination in the insulin-like growth factor 1 null brain is proportionate to its neuronal composition. *J. Neurosci.* **18**, 5673–5681 (1998).
46. Grossfeld, R. M. & Shooter, E. M. A study of the changes in protein composition of mouse brain during ontogenetic development. *J. Neurochem.* **18**, 2265–2277 (1971).

Structure of the human smoothened receptor bound to an antitumour agent

Chong Wang¹, Huixian Wu¹, Vsevolod Katritch¹, Gye Won Han¹, Xi-Ping Huang², Wei Liu¹, Fai Yiu Siu¹, Bryan L. Roth², Vadim Cherezov¹ & Raymond C. Stevens¹

The smoothened (SMO) receptor, a key signal transducer in the hedgehog signalling pathway, is responsible for the maintenance of normal embryonic development and is implicated in carcinogenesis. It is classified as a class frizzled (class F) G-protein-coupled receptor (GPCR), although the canonical hedgehog signalling pathway involves the GLI transcription factors and the sequence similarity with class A GPCRs is less than 10%. Here we report the crystal structure of the transmembrane domain of the human SMO receptor bound to the small-molecule antagonist LY2940680 at 2.5 Å resolution. Although the SMO receptor shares the seven-transmembrane helical fold, most of the conserved motifs for class A GPCRs are absent, and the structure reveals an unusually complex arrangement of long extracellular loops stabilized by four disulphide bonds. The ligand binds at the extracellular end of the seven-transmembrane-helix bundle and forms extensive contacts with the loops.

The SMO receptor is an essential component of the canonical hedgehog signalling pathway, which has a key role in the regulation of embryonic development in animals^{1,2}. In vertebrates, the binding of hedgehog signalling proteins to the receptor patched^{3,4}, a 12-transmembrane protein that inhibits the activity of the SMO receptor⁵, results in translocation of the SMO receptor into cilia⁶. In cilia, the activated SMO receptor induces the modification of GLI transcription factors into the active form, thereby inhibiting the processing of GLI transcription factors into a repressor form. Subsequently, active GLI transcription factors translocate to the nucleus activating the transcription of GLI-targeted genes that control embryonic development and other processes².

The SMO receptor shows a high sequence similarity to the frizzled (FZD) receptors, which mediate the WNT signalling pathway^{7,8}. Both SMO and FZD receptors contain an extracellular domain (ECD) composed of an extracellular cysteine-rich domain (CRD) and an ECD linker domain (Supplementary Fig. 1), a seven-transmembrane helical (7TM) domain and an intracellular carboxy-terminal domain⁹. The CRD of the FZD receptors binds the endogenous lipoglycoprotein ligands, the WNTs^{10,11}; unfortunately, the function of the homologous SMO receptor CRD remains unclear. The 7TM domains of SMO and FZD receptors are reminiscent of class A GPCRs¹² although their sequences are distinct from other GPCRs, sharing less than 10% sequence identity. For this reason and because of their distinct signalling properties, the SMO and FZD receptors have been defined as class F^{9,13}.

The classification of the SMO receptor as a GPCR remains controversial, principally owing to the lack of GPCR-like features in the canonical hedgehog signalling pathway, although emerging evidence indicates that the SMO receptor can share some functional similarities with other classical GPCRs¹⁴. For example, activated SMO receptor can be phosphorylated by a GPCR kinase, leading to β -arrestin translocation and binding¹⁵. Moreover, the SMO receptor can couple to G proteins, particularly G_i¹⁶, which inhibits cAMP production, controlling cell migration¹⁷. Finally, the function of the SMO receptor can be modulated by natural and synthetic small-molecule agonists and antagonists, some of which are potential antitumour agents¹⁸.

Structure of the SMO receptor 7TM domain

An engineered construct of the human SMO receptor with a thermo-stabilized apocytochrome *b*₅₆₂RIL (BRIL) fused to the amino terminus of S190 and the C terminus truncated at Q555, which preserved the ligand-binding property of wild-type human SMO receptor, was expressed, purified and crystallized in complex with the antagonist LY2940680 (refs 19, 20) using a lipidic mesophase method²¹ (Supplementary Figs 2–4). The structure was solved using a 3.5 Å single-wavelength anomalous dispersion (SAD) data set, followed by extending the resolution to 2.5 Å using native data collected from five crystals (Supplementary Table 1).

The SMO receptor structure (Fig. 1) reveals a canonical GPCR 7TM bundle fold with a short helix VIII packed parallel to the membrane bilayer. The ECD linker domain and long extracellular loops (ECLs) form intricate structures stabilized by four disulphide bonds: C193–C213, C217–C295, C314–C390 and C490–C507. The ligand LY2940680 binds in a pocket at the extracellular side of the receptor formed by the 7TM bundle and the ECLs. The receptor crystallizes as a parallel dimer in the crystallographic asymmetric unit (Fig. 1 and Supplementary Fig. 5) with the interface involving helices IV and V, as observed for CXCR4 (ref. 22). It has been reported that the SMO receptor forms a functionally important dimer, although it is unclear whether the crystallographic dimer is the same as that found in the cell membrane²³. Because the difference between the two protomers is small (Supplementary Fig. 5), we will focus on molecule A in the following discussion for brevity, except where otherwise noted.

7TM comparisons with class A GPCRs

Sequence similarity between the SMO receptor and class A GPCRs is very low (less than 10% sequence identity), and SMO and other class F receptors lack most of the conserved class A motifs, including D[E]R^{3,50}Y in helix III, CWXP^{6,50} in helix V and NP^{7,50}XXY in helix VII¹² (superscript numbers reflect the B&W numbering system for GPCRs²⁴, brackets define an alternative residue and X denotes any residue). However, an overlay of the SMO receptor structure with previously solved class A

¹Department of Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. ²National Institute of Mental Health Psychoactive Drug Screening Program, Department of Pharmacology and Division of Chemical Biology and Medicinal Chemistry, University of North Carolina Chapel Hill Medical School, 4072 Genetic Medicine Building, Chapel Hill, North Carolina 27514, USA.

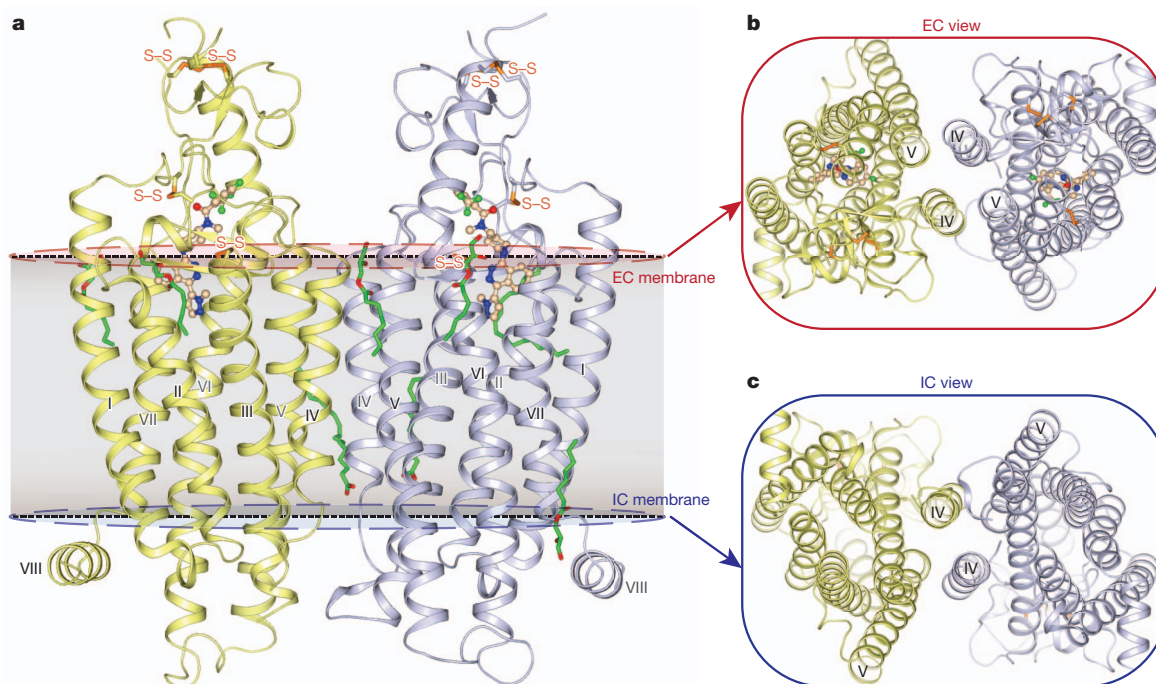


Figure 1 | Overall structure of the human SMO receptor 7TM domain in complex with LY2940680. **a**, Overall structure of the SMO receptor bound to the ligand LY2940680. The SMO receptor is crystallized as a dimer in an asymmetric unit, with molecule A coloured light blue and molecule B

coloured yellow. LY2940680 is shown in wheat carbon. The four disulphide bonds (S–S) are shown in orange sticks. The lipids are shown in sticks with green carbons. Membrane boundaries are shown in dotted lines. **b**, **c**, Extracellular (EC) and intracellular (IC) view of the SMO receptor dimer.

GPCR structures shows relatively high spatial conservation of the 7TM bundle (Fig. 2a, b and Supplementary Fig. 6). Several intracellular structural features of class A GPCR 7TM bundles are also preserved, including a helical turn in short intracellular loop 1 (ICL1) and a short intracellular helix VIII running parallel to the membrane surface, although it has an interface (residues T541, I544 and W545) with helix I (residues T251 and A254) (Fig. 2c) that is distinct from the interface observed in class A GPCRs¹². Structural similarity with class A GPCRs makes transplanting the B&W numbering²⁴ system to class F receptors possible on the basis of structural superposition. In each helix, the following residues are assigned number 50: T245^{1.50}, F274^{2.50}, W339^{3.50}, W365^{4.50}, V411^{5.50}, S468^{6.50} and I530^{7.50} (Supplementary Figs 1, 2 and 6). The numbering of the other residues in each helix is counted relative to the X.50 position according to the B&W numbering system²⁴.

Despite the overall structural conservation, the 7TM fold of the SMO receptor has many distinct features. For example, when compared with class A GPCRs the extracellular tip of helix V is shifted towards the ligand-binding cavity. Most importantly helices V, VI and VII of the SMO receptor lack the most conserved prolines (P^{5.50}, P^{6.50} and P^{7.50}), which have pivotal roles in the activation process of class A GPCRs¹². In the β_2 adrenergic receptor, P^{5.50} has been shown to act as a local trigger of GPCR activation along with I^{3.40} and F^{6.44} (ref. 25). Instead of P^{5.50} in helix V of class A GPCRs, the SMO receptor has P407^{5.46} in an adjacent helical turn (Fig. 2d). In helix VI, P^{6.50} induces a kink in class A GPCRs, which facilitates the large movement of the intracellular segment of helix VI during activation. The SMO receptor has no proline in helix VI, and thus this helix is straighter than in class A GPCRs (Fig. 2e). Similarly, in helix VII, which typically has the conserved NPXXY motif in class A GPCRs, the proline is also absent (Fig. 2f). Although these prolines are missing in the SMO receptor, we observed a large number of glycines in helices V, VI and VII (Supplementary Fig. 2). Conceivably, these glycines could facilitate both flexibility and bending of the helices, thereby enabling 7TM packing and conformational changes during the activation of the SMO receptor. In the current structure of the SMO receptor in complex with an antagonist, helix

VI is found in an inactive-like, closed state (Supplementary Fig. 7), presumably precluding G-protein binding.

Binding site of LY2940680

LY2940680 is a SMO receptor antagonist designed for the treatment of solid tumours¹⁹. The SMO receptor binding pocket has a long and narrow shape and is connected to the extracellular aqueous environment through a small opening formed by the ECD linker domain, ECL2 and ECL3 (Fig. 3a). This orifice probably facilitates small-molecule ligand entry into the 7TM core region. Residues from the extracellular tips of helices I, II, V, and VII interact with LY2940680, most notably R400^{5.39} of helix V, which hydrogen bonds with the phthalazine ring system of the ligand. Most of the other contact residues belong to the ECD linker domain and ECLs (Fig. 3b, c and Supplementary Fig. 8). Several structured water molecules are identified in the ligand pocket, including two waters mediating the hydrogen-bonding network between R400^{5.39}, H470^{6.52}, D473^{6.55}, E518^{7.38} and N521^{7.41} side chains (Supplementary Fig. 9). Although these waters do not directly contact LY2940680, they may have an important role in the conformational properties and dynamics of the pocket. Mutation of D473^{6.55}—which participates in this water-mediated hydrogen-bonding network—to histidine results in resistance to the approved Genentech drug GDC-0449 (ref. 26). Direct contact of this residue with LY2940680 is limited (distance between the carboxylate of D473^{6.55} and LY2940680 is 4.04 Å in molecule A and 4.31 Å in molecule B). Cyclopamine, a naturally occurring steroid and the first identified small-molecule SMO receptor ligand, inhibits the hedgehog signalling pathway²⁷. Radioligand assays revealed that LY2940680 and the SMO receptor agonist SAG compete with the binding of ³H-cyclopamine (Supplementary Fig. 3), indicating that these ligands bind within the long and narrow cavity embedded in the 7TM domain of the SMO receptor.

ECD linker domain and ECL structures

The SMO receptor has a unique ECD linker domain, and its ECLs are long compared to most class A GPCRs. These extracellular domains

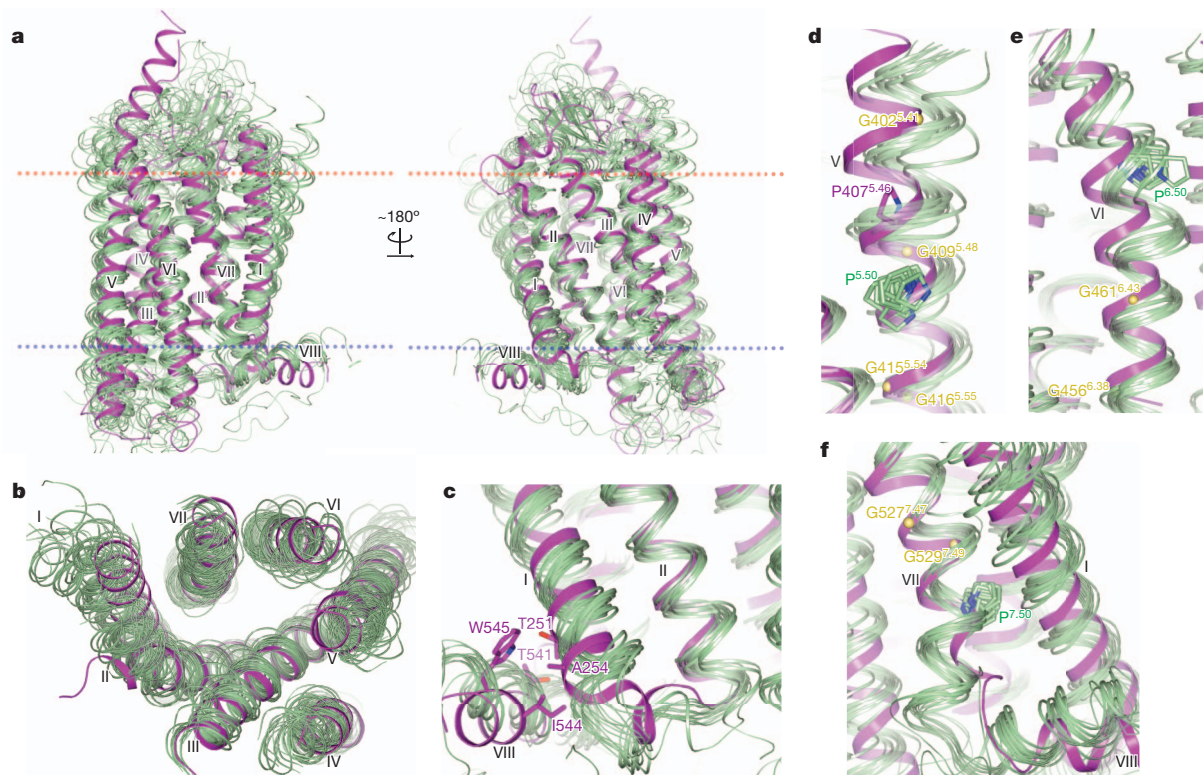


Figure 2 | Comparison of the 7TM bundle of the SMO receptor with class A GPCRs. **a**, Side view of superimposed structures of the SMO receptor (magenta) and class A GPCRs (light green). **b**, Extracellular view of the superimposed structures. **c**, Superposition of ICL1 and helix VIII. The residues involved in the packing interface of helix VIII and helix I in the SMO receptor are shown in sticks with magenta carbons. **d–f**, Superposition of helices V–VIII.

are organized into complex tertiary structures through covalent and noncovalent interactions forming a lid on the 7TM bundle (Fig. 4a, b). The unusually long ECL1 (Fig. 4c) is connected to the ECD linker domain through a disulphide bond between C217 and C295, which divides ECL1 into two distinctive segments. Preceding C295, there is a short α -helical structure (G288 to V294) stabilized by cation- π interactions between the guanidinium group of R290 and the imidazole group of H231^{1,36} from helix I. After C295, ECL1 is packed into a U-shaped loop that is stabilized by an ionic interaction between R302 and E305. In addition, this loop segment forms contacts with the ECD linker domain through a hydrogen-bonding network involving residues N202, S205, D298 and T300. ECL2 forms a β -hairpin and connects to helix III by a disulphide bond between C314^{3,25} and C390. This loop is positioned deep within the cavity formed by the 7TM bundle and makes extensive contacts with LY2940680. ECL3 is the longest loop of the SMO receptor and forms a protrusion from the 7TM bundle into the extracellular space (Fig. 4d). The long extension of helix VI adopts a well-ordered α -helical structure that is partially stabilized by an ionic interaction between E479 and R482. This α -helical extension connects to helix VI through a 45° non-proline kink that is stabilized by several water molecules (Supplementary Fig. 10). On top of the ECL3 helical structure there is C490, which forms a disulphide bond with C507. The loop between C490 and C507 is mostly disordered, whereas the segment between C507 and the extracellular tip of helix VII forms an extended strand. ECL3 also makes contact with the ECD linker domain: R485 interacts with E208 within the ECD linker domain through a salt bridge; the amide side chain of Q491 and the guanidinium group of R512 form hydrogen bonds with the main-chain carbonyl groups of V195 and L221, respectively. The integrity of the ECL structures is essential for maintaining the SMO receptor in an

The P^{5.50}, P^{6.50} and P^{7.50} residues conserved in class A GPCRs are shown in sticks. In **d**, P407^{5.46} in the SMO receptor is shown in sticks. The α carbons of glycines in the SMO receptor are shown as yellow spheres. Structures of class A GPCRs used were (PDB accessions): 1U19, 2RH1, 2YCW, 3RZE, 3PBL, 3UON, 4DAJ, 3EML, 3V2W, 3ODU, 4DJH, 4EA3, 4DKL, 4EJ4, 3VW7 and 4GRV.

inactive state, as disruption of the extracellular structures by mutations of the extracellular cysteines increases SMO receptor activity²⁸.

In the extracellular region, the only structural feature that the SMO receptor shares with class A GPCRs is the β -hairpin structure of ECL2 that is linked to the extracellular tip of helix III through a disulphide bond (Fig. 5). The corresponding cysteine in position 3.25 (B&W numbering²⁴), is conserved in nearly all class A and other GPCRs. The β -hairpin structure of ECL2 appears to be a hallmark of class A peptide-binding GPCRs^{22,29–34}, which has been shown by docking studies³¹ and peptide-receptor co-crystal structures^{22,32} to have an important role in the recognition of peptide ligands. The ECL2s of peptide-binding receptors all point outwards from the 7TM core domain, leaving relatively open and spacious binding cavities for their cognate peptide ligands (Fig. 5c–f). By contrast, the β -hairpin structure of ECL2 in rhodopsin folds on top of its covalently attached ligand retinal, sealing the extracellular entrance of the pocket (Fig. 5b). Interestingly, the ECL2 structure of the SMO receptor is distinct from both rhodopsin and peptide-binding GPCRs. Although ECL2 sits much deeper in the SMO receptor than in class A peptide receptors and occupies a substantial space in the cavity of 7TM bundle (Fig. 5a), unlike in rhodopsin, the β -hairpin in the SMO receptor does not occlude the ligand entrance. Instead, the ECL2 of the SMO receptor is positioned laterally to LY2940680 (Supplementary Fig. 11), forming a large part of the binding pocket for this ligand.

Homology with frizzled family receptors

Within class F receptors, one can find a gapless alignment at trans-membrane helices (Supplementary Fig. 1), and 45 residues are fully conserved within the ECD linker and 7TM domains (Fig. 6a, b). The cysteines that form disulphide bonds maintaining the structures of the ECD linker domain and the ECLs in the SMO receptor are highly

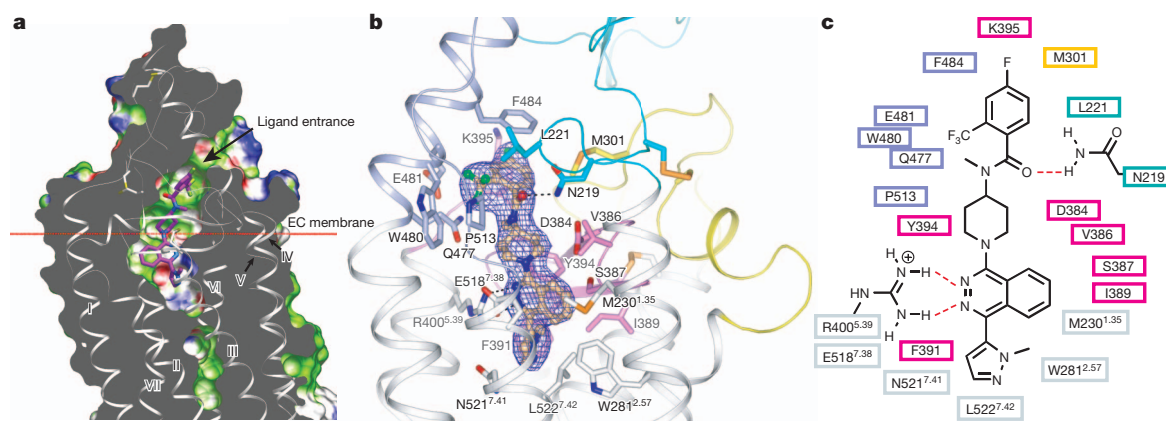


Figure 3 | Ligand-binding pocket for LY2940680. **a**, The ligand-binding pocket surface is coloured according to binding properties (green, hydrophobic; blue, hydrogen-bond donor; red, hydrogen-bond acceptor). LY2940680 is shown by sticks with magenta carbons. The membrane boundary is shown as a red line. **b**, $2F_o - F_c$ map (blue mesh) is shown for LY2940680 (contoured at 1.0σ , $0.10 \text{ e } \text{\AA}^{-3}$). Binding-pocket residues (4.0 \AA cutoff distance

from LY2940680) are shown in sticks with colours indicating the location (white, transmembrane helices; cyan, ECD linker domain; yellow, ECL1; pink, ECL2; light blue, ECL3). The hydrogen-bond interactions between R400^{5.39}, N219 and LY2940680 are shown as black dashed lines. **c**, Diagram of ligand interactions in the binding pocket. The colour scheme of the boxes for the residues is the same as in **b**.

preserved in the FZD receptors (the only exception is FZD4, in which the disulphide bond in ECL3 is missing, probably owing to a very short loop), indicating the importance of disulphide bonds in maintaining the ECL structures for the FZD receptors. In the extracellular half of the 7TM bundle, the conserved residues form a cluster of hydrophobic side chains buried between helices III (F318^{3.29}, Y322^{3.33}, M326^{3.37}), V (F403^{5.42}, V404^{5.43}, P407^{5.46}) and ECL2 (V392), which is apparently important for the structural integrity of these receptors. Closer to the intracellular membrane boundary, there is an unusually high number of conserved tryptophans: W331^{3.42} and W339^{3.50} in helix III, W365^{4.50} in helix IV and W535^{7.55} in helix VII. The latter

tryptophan, conserved among class F and superimposable with the location of the NPXXY motif of class A, is shown to have an important role in receptor activation, as mutation of W535^{7.55} leads to a constitutively active SMO receptor³⁵. In the FZD receptors, the KTXXXW motif in helix VIII is highly conserved and has been shown to be critical for the activation of the WNT/ β -catenin signalling pathway by interacting directly with dishevelled^{36,37}. Sequence alignment shows that the SMO receptor has an extra alanine between lysine and threonine, but that the rest of the motif is conserved. In the SMO receptor structure, this motif stabilizes the α -helical structure of helix VIII, which packs parallel to the membrane (Fig. 6c). The hydroxyl group of the conserved

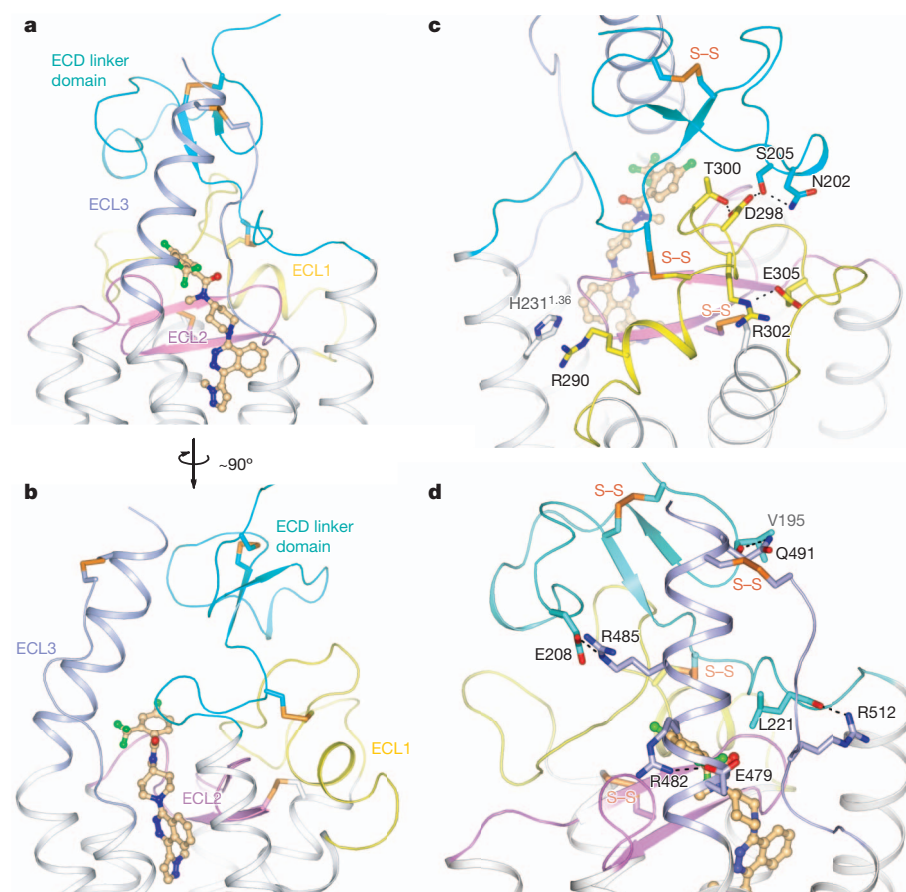


Figure 4 | The structure of the ECD linker domain and ECLs.

a, **b**, Different views of the cartoon presentation of the structure of the ECD linker domain and ECLs. ECD linker domain, cyan; ECL1, yellow; ECL2, pink; ECL3, light blue. Disulphide bonds are shown in orange sticks. **c**, Side-chain interactions that stabilize the structure of ECL1. The hydrogen-bond network among residues N202, S205, D298 and T300 and the ionic interaction between R302 and E305 are shown in dashed lines. **d**, Side-chain interactions that stabilize the structure of ELC3. The ionic interactions between E208 and R485, E479 and R482, and hydrogen-bond interactions between Q491 and V195, R512 and L221 are shown as dashed lines.

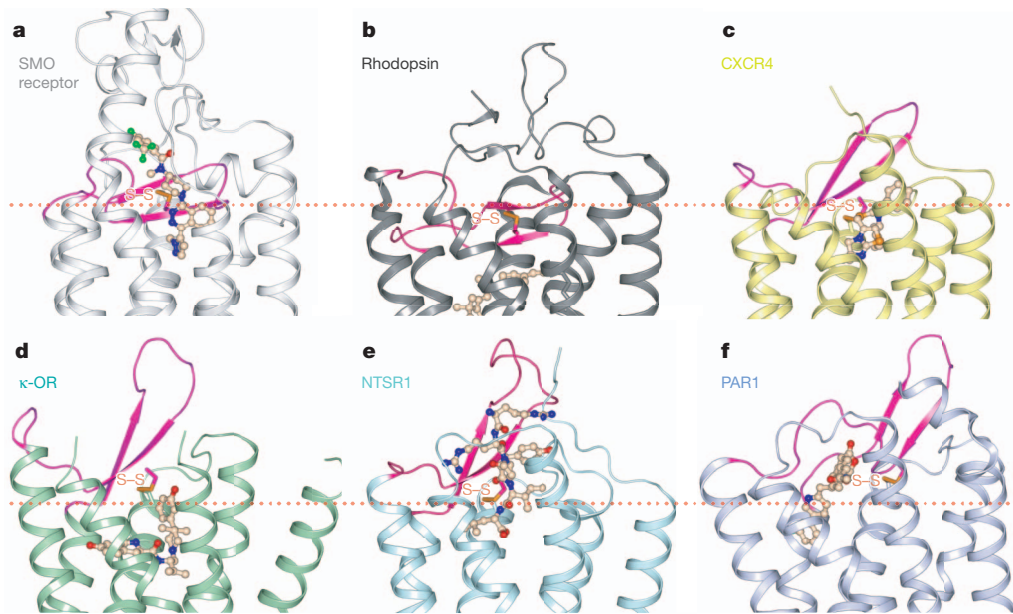


Figure 5 | Comparison of ECL2 of the SMO receptor with class A GPCRs. **a–f**, Cartoon presentation of the extracellular part of the structures for the SMO receptor (white; **a**), rhodopsin (PDB 1U19) (dark grey; **b**), CXCR4 chemokine receptor (PDB 3ODU) (yellow; **c**), κ -opioid receptor (PDB 4DJH) (light green, representing the opioid receptor family; **d**), neurotensin receptor

NTSR1 (PDB 4GRV) (cyan; **e**), and protease-activated receptor 1 (PAR1) (PDB 3VW7) (light blue; **f**). The ECL2 of each structure is shown in magenta. All ligands are shown in wheat. The approximate position of the extracellular membrane boundary is shown in orange dotted lines. The disulphide bonds connecting ECL2 and helix III are shown in orange sticks.

T541 of this motif forms a hydrogen bond with the main-chain carbonyl group of V536 at the intracellular end of helix VII, whereas the indole nitrogen of W545 forms a hydrogen-bond interaction with the hydroxyl group of T251^{1,56}, which is conserved in the FZD receptors. The α -helical structure of helix VIII therefore probably has a critical role for the interaction of activated FZD receptors with downstream signalling proteins, such as dishevelled.

Conclusion

Evolutionarily, class F receptors precede class A receptors, as revealed by extensive phylogenetic analysis³⁸. Despite the earlier emergence, the diversity within class F is considerably less than that of class A. This remarkable conservation probably reflects the pivotal function of the SMO and FZD receptors in the regulation of cell proliferation polarity and differentiation along with tissue formation, some of the

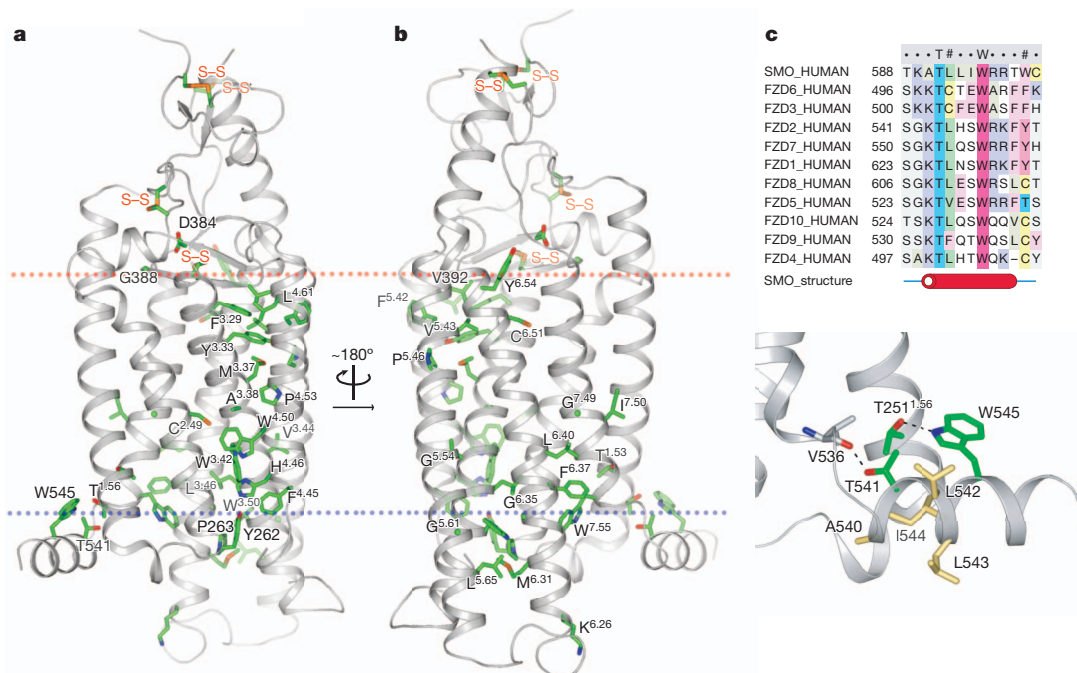


Figure 6 | Structural insight for the FZD receptors on the basis of their sequence homology with the SMO receptor. **a, b**, Fully conserved residues in class F members are highlighted on the SMO structure. The side chains of all these conserved residues are shown in sticks with green carbons. **c**, Alignment of the KTXXXW motif in the FZD receptors to helix VIII of the SMO receptor. The similar α -helical structure could exist for the KTXXXW

motif of the FZD receptors, which is stabilized by the hydrogen-bond interactions between the hydroxyl group of the conserved residue T541 and the main-chain carbonyl group of V536, the indole nitrogen of the conserved residue W545 and the hydroxyl group of the conserved residue T251^{1,56}, shown in dashed lines. The non-conserved residues are coloured yellow.

most fundamental physiological processes for multicellular metazoan. The SMO receptor structure highlights the incredible use of the 7TM structural fold, with little sequence similarity to other GPCR classes, and is a great example of structural conservation in protein space. As we learn more about different GPCR classes and their structure–function relationships, including the expansion of many different intracellular interacting partners beyond G proteins, it is probable that the term ‘GPCR’ may be of limited value in describing the incredible power of the 7TM fold³⁹ in biology.

METHODS SUMMARY

The BRIL-ΔCRD-SMO-ΔC construct for structural study was expressed in *Spodoptera frugiperda* (Sf9) cells. Ligand-binding and functional assays were performed as described in Methods. Sf9 cell membranes were solubilized using 1% (w/v) *n*-dodecyl-β-D-maltopyranoside and 0.2% (w/v) cholesteryl hemisuccinate, and purified by immobilized metal ion affinity chromatography (IMAC), followed by reverse IMAC after cleaving N-terminal Flag–10×His tags by His-tagged tobacco etch virus protease. The purified protein at a concentration of 50–60 mg ml^{−1} was mixed with monoolein and cholesterol in a ratio of 40%:54%:6% (w/w/w) to form lipidic cubic phase (LCP), from which the receptor was crystallized. Crystals were grown at 20 °C in 40 nl protein-laden LCP boluses overlaid by 800 nl of precipitant solutions as described in Methods. Crystals were collected from the LCP matrix and flash frozen in liquid nitrogen. X-ray diffraction data were collected on the 23ID-D beamline at the Advanced Photon Source using a 20 μm minibeam at a wavelength of 1.0330 Å. Phase information was obtained by SAD data obtained from a single crystal soaked in tantalum bromide. Data collection, processing, structure solution and refinement are described in Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 15 January; accepted 12 April 2013.

Published online 1 May 2013.

- Ingham, P. W. & McMahon, A. P. Hedgehog signaling in animal development: paradigms and principles. *Genes Dev.* **15**, 3059–3087 (2001).
- Robbins, D. J., Fei, D. L. & Riobo, N. A. The Hedgehog signal transduction network. *Sci. Signal.* **5**, re6 (2012).
- Marigo, V., Davey, R. A., Zuo, Y., Cunningham, J. M. & Tabin, C. J. Biochemical evidence that patched is the Hedgehog receptor. *Nature* **384**, 176–179 (1996).
- Stone, D. M. *et al.* The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. *Nature* **384**, 129–134 (1996).
- Taipale, J., Cooper, M. K., Maiti, T. & Beachy, P. A. Patched acts catalytically to suppress the activity of Smoothened. *Nature* **418**, 892–896 (2002).
- Corbit, K. C. *et al.* Vertebrate Smoothened functions at the primary cilium. *Nature* **437**, 1018–1021 (2005).
- Klaus, A. & Birchmeier, W. Wnt signalling and its impact on development and cancer. *Nature Rev. Cancer* **8**, 387–398 (2008).
- Huang, H. C. & Klein, P. S. The Frizzled family: receptors for multiple signal transduction pathways. *Genome Biol.* **5**, 234 (2004).
- Schulte, G. International Union of Basic and Clinical Pharmacology. LXXX. The class Frizzled receptors. *Pharmacol. Rev.* **62**, 632–667 (2010).
- Bhanot, P. *et al.* A new member of the frizzled family from *Drosophila* functions as a Wingless receptor. *Nature* **382**, 225–230 (1996).
- Janda, C. Y., Waghray, D., Levin, A. M., Thomas, C. & Garcia, K. C. Structural basis of Wnt recognition by Frizzled. *Science* **337**, 59–64 (2012).
- Katritch, V., Cherezov, V. & Stevens, R. C. Structure-function of the G protein-coupled receptor superfamily. *Annu. Rev. Pharmacol. Toxicol.* **53**, 531–556 (2013).
- Foord, S. M. *et al.* International Union of Pharmacology. XLVI. G protein-coupled receptor list. *Pharmacol. Rev.* **57**, 279–288 (2005).
- Ayers, K. L. & Therond, P. P. Evaluating Smoothened as a G-protein-coupled receptor for Hedgehog signalling. *Trends Cell Biol.* **20**, 287–298 (2010).
- Chen, W. *et al.* Activity-dependent internalization of Smoothened mediated by β-arrestin 2 and GRK2. *Science* **306**, 2257–2260 (2004).
- Riobo, N. A., Saucy, B., Dilizio, C. & Manning, D. R. Activation of heterotrimeric G proteins by Smoothened. *Proc. Natl Acad. Sci. USA* **103**, 12607–12612 (2006).
- Polizio, A. H. *et al.* Heterotrimeric Gi proteins link Hedgehog signaling to activation of Rho small GTPases to promote fibroblast migration. *J. Biol. Chem.* **286**, 19589–19596 (2011).
- Heretsch, P., Tzagkaroulaki, L. & Giannis, A. Modulators of the hedgehog signaling pathway. *Bioorg. Med. Chem.* **18**, 6613–6624 (2010).
- Bender, M. H. *et al.* Identification and characterization of a novel smoothened antagonist for the treatment of cancer with deregulated hedgehog signaling. *Cancer Res.* **71**, 2819 (2011).
- Hipskind, P. A., Patel, B. K. & Wilson, T. Disubstituted phthalazine hedgehog pathway antagonists. US patent 8,273,742 B2 (2010).
- Cherezov, V. *et al.* High-resolution crystal structure of an engineered human β₂-adrenergic G protein-coupled receptor. *Science* **318**, 1258–1265 (2007).
- Wu, B. *et al.* Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **330**, 1066–1071 (2010).
- Zhao, Y., Tong, C. & Jiang, J. Hedgehog regulates smoothened activity by inducing a conformational switch. *Nature* **450**, 252–258 (2007).
- Ballesteros, J. A. & Weinstein, H. in *Methods in Neurosciences* Vol. 25 (ed. Sealfon, S. C.) 366–428 (Academic, 1995).
- Rasmussen, S. G. *et al.* Structure of a nanobody-stabilized active state of the β₂ adrenoceptor. *Nature* **469**, 175–180 (2011).
- Yauch, R. L. *et al.* Smoothened mutation confers resistance to a Hedgehog pathway inhibitor in medulloblastoma. *Science* **326**, 572–574 (2009).
- Taipale, J. *et al.* Effects of oncogenic mutations in *Smoothened* and *Patched* can be reversed by cyclopamine. *Nature* **406**, 1005–1009 (2000).
- Carroll, C. E., Marada, S., Stewart, D. P., Ouyang, J. X. & Ogden, S. K. The extracellular loops of Smoothened play a regulatory role in control of Hedgehog pathway activation. *Development* **139**, 612–621 (2012).
- Granier, S. *et al.* Structure of the δ-opioid receptor bound to naltrindole. *Nature* **485**, 400–404 (2012).
- Manglik, A. *et al.* Crystal structure of the μ-opioid receptor bound to a morphinan antagonist. *Nature* **485**, 321–326 (2012).
- Thompson, A. A. *et al.* Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* **485**, 395–399 (2012).
- White, J. F. *et al.* Structure of the agonist-bound neurotensin receptor. *Nature* **490**, 508–513 (2012).
- Wu, H. *et al.* Structure of the human κ-opioid receptor in complex with JDTic. *Nature* **485**, 327–332 (2012).
- Zhang, C. *et al.* High-resolution crystal structure of human protease-activated receptor 1. *Nature* **492**, 387–392 (2012).
- Xie, J. *et al.* Activating *Smoothened* mutations in sporadic basal-cell carcinoma. *Nature* **391**, 90–92 (1998).
- Umbhauer, M. *et al.* The C-terminal cytoplasmic Lys-Thr-X-X-X-Trp motif in frizzled receptors mediates Wnt/β-catenin signalling. *EMBO J.* **19**, 4944–4954 (2000).
- Wong, H. C. *et al.* Direct binding of the PDZ domain of Dishevelled to a conserved internal sequence in the C-terminal region of Frizzled. *Mol. Cell* **12**, 1251–1260 (2003).
- Krishnan, A., Almen, M. S., Fredriksson, R. & Schiöth, H. B. The origin of GPCRs: identification of mammalian like *Rhodopsin*, *Adhesion*, *Glutamate* and *Frizzled* GPCRs in fungi. *PLoS ONE* **7**, e29817 (2012).
- Schwartz, T. W. & Rosenkilde, M. M. Is there a ‘lock’ for all agonist ‘keys’ in 7TM receptors? *Trends Pharmacol. Sci.* **17**, 213–216 (1996).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the National Institutes of Health Common Fund grant P50 GM073197 for technology development (V.C. and R.C.S.), PSI:Biologics grant U54 GM094618 for biological studies and structure production (target GPCR-131) (V.K., V.C. and R.C.S.); F32 DK088392 (F.Y.S.); R01 MH61887, U19 MH82441, R01 DA27170 and the NIMH Psychoactive Drug Screening Program (X.-P.H. and B.L.R.) and the Michael Hooker Chair of Pharmacology (B.L.R.). We thank J. Velasquez for help with molecular biology, T. Trinh and M. Chu for help with baculovirus expression, K. Kadyshchinskaya for assistance with figure preparation, A. Walker for assistance with manuscript preparation, D. Wacker for assistance with SAD data collection and processing and J. Smith, R. Fischetti and N. Sanishvili for assistance in development and use of the minibeam and beamtime at beamline 23-ID at the Advanced Photon Source, which is supported by National Cancer Institute grant Y1-CO-1020 and National Institute of General Medical Sciences grant Y1-GM-1104.

Author Contributions C.W. designed and made the constructs, purified and crystallized the receptor in LCP, optimized crystallization conditions, grew crystals for data collection, assisted with crystal collection and heavy-atom soaking experiment, collected the diffraction data and prepared the manuscript. H.W. performed baculovirus expression of the receptor, purified the receptor, optimized crystallization conditions and prepared the manuscript. V.K. prepared the manuscript. G.W.H. solved and refined the structure and assisted with preparation of the manuscript. X.-P.H. performed the radioligand binding assay. W.L. assisted with data collection. F.Y.S. provided material and conditions for heavy-atom soaking experiment. B.L.R. supervised the radioligand binding experiment and assisted with preparing the manuscript. V.C. collected crystals, performed the heavy-atom soaking experiment, collected and processed diffraction data and assisted with preparation of the manuscript. R.C.S. was responsible for the overall project strategy and management and wrote the manuscript.

Author Information The coordinates and the structure factors have been deposited in the Protein Data Bank under the accession code 4JKV. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.C.S. (stevens@scripps.edu).

METHODS

Generation of BRIL-ACRD-SMO-AC fusion construct for structural studies.

Human SMO gene was obtained from Origene. A thermally stabilized apocytochrome *b*₅₆₂ RIL from *E. coli* (M7W, H102L, R106L), referred to as BRIL, was fused to the truncated N terminus at S190 of human SMO receptor, using overlapping PCR. The C terminus of SMO receptor was truncated at Q555. The resulting receptor chimera sequence was subcloned into a modified pFastBac1 vector (Invitrogen), designated as pFastBac1-833100, which contained an expression cassette with a haemagglutinin (HA) signal sequence followed by a Flag tag, a 10×His tag, and a tobacco etch virus (TEV) protease recognition site at the N terminus before the receptor sequence. Subcloning into the pFastBac1-833100 was achieved using PCR with primer pairs encoding restriction sites KpnI at the 5' and HindIII at the 3' termini with subsequent ligation into the corresponding restriction sites found in the vector.

Expression and purification of BRIL-ACRD-SMO-AC protein for crystallization. The resulting BRIL-ACRD-SMO-AC construct was expressed in *Spodoptera frugiperda* (Sf9) insect cells using the Bac-to-Bac Baculovirus Expression System (Invitrogen). Sf9 cells at cell density of $2-3 \times 10^6$ cells per ml were infected with baculovirus at 27 °C. Cells were collected by centrifugation at 48 h after infection and stored at -80 °C until use.

Insect cell membranes were lysed by thawing frozen cell pellets in a hypotonic buffer containing 10 mM HEPES, pH 7.5, 10 mM MgCl₂, 20 mM KCl and EDTA-free complete protease inhibitor cocktail tablets (Roche). Extensive washing of the raw membranes was performed by repeated centrifugation (two to three times) in a high osmotic buffer comprised of 1.0 M NaCl in the hypotonic buffer described above.

The washed membranes were re-suspended into buffer containing 30 μM LY2940680 (Active Biochemicals Co.), 2 mg ml⁻¹ iodoacetamide (Sigma) and EDTA-free complete protease inhibitor cocktail tablets, and incubated at 4 °C for 1 h before solubilization. The membranes were then solubilized in buffer containing 50 mM HEPES, pH 7.5, 200 mM NaCl, 1% (w/v) *n*-dodecyl-β-D-maltopyranoside (DDM; Anatrace) and 0.2% (w/v) cholesteryl hemisuccinate (CHS, Sigma), for 3–4 h at 4 °C. The supernatant containing solubilized SMO protein was isolated from the cell debris by high-speed centrifugation, and subsequently incubated with TALON IMAC resin (Clontech) overnight at 4 °C in the presence of 20 mM imidazole and 1 M NaCl. After binding, the resin was washed with 10-column volumes of wash I buffer comprised of 50 mM HEPES, pH 7.5, 800 mM NaCl, 10% (v/v) glycerol, 0.1% (w/v) DDM, 0.02% (w/v) CHS, 8 mM ATP, 20 mM imidazole, 10 mM MgCl₂ and 15 μM LY2940680, followed by 6-column volumes of wash II buffer comprised of 50 mM HEPES, pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, 0.05% (w/v) DDM, 0.01% (w/v) CHS, 50 mM imidazole and 20 μM LY2940680. The protein was then eluted by 3-column volumes of elution buffer containing 50 mM HEPES, pH 7.5, 300 mM NaCl, 10% (v/v) glycerol, 0.03% (w/v) DDM, 0.006% (w/v) CHS, 250 mM imidazole and 50 μM LY2940680. PD MiniTrap G-25 column (GE Healthcare) was used to remove imidazole. The protein was then treated overnight with TEV protease (His-tagged) to cleave the N-terminal His tag and FLAG tag. TEV protease and cleaved N-terminal fragment were removed by TALON IMAC resin incubation at 4 °C for 2 h. The tag-less protein was collected as the TALON IMAC column flow-through. The protein was then concentrated to 50–60 mg ml⁻¹ with a 100 kDa cutoff Vivaspinn concentrator. Protein monodispersity was tested by analytical size-exclusion chromatography (aSEC). Typically, the aSEC profile showed a monodisperse peak.

Lipidic cubic phase crystallization. Protein samples of the SMO receptor in a complex with LY2940680 were reconstituted into lipidic cubic phase (LCP) by mixing with molten lipid (10% (w/w) cholesterol, 90% (w/w) monolein) in a mechanical syringe mixer⁴⁰ at a ratio of 2/3 (v/v) protein solution/lipid. LCP crystallization trials were performed using an NT8-LCP crystallization robot (Formulatrix) as previously described⁴¹. 96-well glass sandwich plates (Marienfeld) were incubated and imaged at 20 °C using an automated incubator/imager (RockImager 1000, Formulatrix). Initial crystal hits were found from a precipitant condition containing 100 mM HEPES, pH 7.4, 30% (v/v) PEG400 and 100 mM ammonium fluoride. After optimization, crystals grew in 100 mM HEPES, pH 7.8, 70 mM ammonium fluoride, 32% (v/v) PEG400 and 4–8% (v/v) polypropylene glycol P 400 to the size of 100 × 20 × 20 μm for 2–3 days, and were collected using MiTeGen micromounts and flash-frozen in liquid nitrogen for data collection.

Crystallographic data collection and processing. X-ray data were collected at the 23ID-D beamline at the Advanced Photon Source using a 20 μm minibeam at a wavelength of 1.0330 Å and a MarMosaic 300 CCD detector. Crystals were aligned and data were collected using strategy similar to other GPCR structures⁴². Typically 20 frames at 1° oscillation and 1-s exposure with non-attenuated beam were collected following by a translation of the crystal to a non-exposed position

or changing the crystal to minimize the effect of radiation damage. A complete data set was obtained by indexing, integrating, scaling and merging data from five crystals using HKL2000 (ref. 43) (Supplementary Table 1).

Experimental phasing. The attempts to find a molecular replacement solution using all previous class A GPCR structures as a search model did not generate any reliable solutions due to the low sequence similarity. Therefore, experimental phasing was performed by soaking the crystals in the presence of 5 mM tantalum bromide ([Ta₆Br₁₂]²⁺•2Br⁻, Jena Bioscience) for 24 h. The data were collected at 23ID-D beamline at the Advanced Photon Source using the peak wavelength of the tantalum L3 edge (9.880 keV). A complete 360° data set was acquired from a single crystal by using a 20 μm minibeam at 50 × attenuation with 1° oscillation and 1-s exposure per frame and collecting 30° wedges with direct and inverse beam. The SAD data set was integrated and scaled at 3.5 Å resolution using HKL2000 (ref. 43), and PHENIX.AutoSol⁴⁴ was used to search for the heavy-atom sites.

Structure determination and refinement. The structure was initially solved using 3.5 Å SAD data, collected from a crystal soaked with tantalum bromide cluster, with PHENIX.AutoSol⁴⁴, with the map clearly showing transmembrane helices. Further heavy-atom refinement and phasing, combining the high-resolution native data and SAD (SIRAS), was carried out using SHARP⁴⁵ based on two heavy-atom sites identified from the anomalous difference map. Density modification and automatic tracing were then performed using PHENIX.AutoBuild⁴⁶. Refinement was performed by rounds of REFMAC5 (ref. 47) and autoBUSTER (v.2.8.0, Global Phasing Ltd) using the 2.5 Å resolution native data set followed by manual examination and rebuilding of the refined coordinates in the program Coot⁴⁸ using both 2F_o - F_c and F_o - F_c maps, as well as omit maps. Data collection and refinement statistics are shown in Supplementary Table 1.

Radioligand binding assays. Radioligand binding assays used Sf9 pellets expressing the crystallization construct BRIL-ACRD-SMO-AC (described in expression section) and crude HEK 293T membrane preparations expressing wild-type human SMO receptor in 96-well plates at a final volume of 125 μl. To obtain crude HEK 293T membrane preparations, HEK 293T cells were transfected with a human SMO receptor expression plasmid for 24 h and scraped into conical centrifuge tubes. Collected cells were centrifuged at 1,000g for 10 min and the cell pellet was hypotonically lysed by cold lysis buffer (50 mM Tris-HCl, pH 7.4). Crude membrane fractions were isolated by centrifugation at 21,000g for 20 min at 4 °C. The membrane pellets were re-suspended with lysis buffer at 3× volume of pellet size, subjected to protein concentration determination, and stored in aliquots at -80 °C if not used immediately. To determine equilibrium dissociation constant (K_d) for ³H-cyclopamine, a series of eight concentrations of ³H-cyclopamine (0.2–36 nM in triplicate) were incubated with 6 μg SMO Sf9 membranes or 20 μg of above SMO HEK 293T membranes in binding buffer (50 mM HEPES, 3 mM MgCl₂, EDTA-free protease inhibitor cocktail and 0.5 mg ml⁻¹ BSA, pH 7.2, modified from ref. 49) for 2.5 h in the dark at room temperature (20 °C). Nonspecific binding was defined by 10 μM SMO receptor antagonist LY2940680. To determine equilibrium dissociation constant (K_i) for cyclopamine, LY2940680 and SAG (Cayman Chemical), a series of 11 concentrations of test compound (0.1 nM to 10 μM in triplicate sets) were incubated with a fixed concentration of ³H-cyclopamine (~K_d of ³H-cyclopamine) and SMO Sf9 membranes or SMO HEK293 T membranes for 2.5 h in the dark and at room temperature. At the end of the incubation period, the reactions were stopped by rapid filtration onto 0.3% PEI-soaked GF/A filters and washed three times with cold PBS (pH 7.2). The filters were then microwave dried and scintillant was melted on the filters on a hot plate. The filters were wrapped in plastic wrap and counted for radioactivity. Results (Supplementary Fig. 3) were analysed using GraphPad Prism 5.0.

- Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols* **4**, 706–731 (2009).
- Cherezov, V., Peddi, A., Muthusubramanian, L., Zheng, Y. F. & Caffrey, M. A robotic system for crystallizing membrane and soluble proteins in lipidic mesophases. *Acta Crystallogr. D* **60**, 1795–1807 (2004).
- Cherezov, V. et al. Rastering strategy for screening and centring of microcrystal samples of human membrane proteins with a sub-10 μm size X-ray synchrotron beam. *J. R. Soc. Interface* **6** (suppl. 5), S587–S597 (2009).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Terwilliger, T. C. et al. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D* **65**, 582–601 (2009).
- Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59**, 2023–2030 (2003).

46. Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the *PHENIX AutoBuild* wizard. *Acta Crystallogr. D* **64**, 61–69 (2008).
47. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
48. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallogr. D* **66**, 486–501 (2010).
49. Rominger, C. M. *et al.* Evidence for allosteric interactions of antagonist binding to the Smoothed receptor. *J. Pharmacol. Exp. Ther.* **329**, 995–1005 (2009).

Atmospheric confinement of jet streams on Uranus and Neptune

Yohai Kaspi¹, Adam P. Showman², William B. Hubbard², Oded Aharonson^{1,3} & Ravit Helled⁴

The observed cloud-level atmospheric circulation on the outer planets of the Solar System is dominated by strong east–west jet streams. The depth of these winds is a crucial unknown in constraining their overall dynamics, energetics and internal structures. There are two approaches to explaining the existence of these strong winds. The first suggests that the jets are driven by shallow atmospheric processes near the surface^{1–3}, whereas the second suggests that the atmospheric dynamics extend deeply into the planetary interiors^{4,5}. Here we report that on Uranus and Neptune the depth of the atmospheric dynamics can be revealed by the planets' respective gravity fields. We show that the measured fourth-order gravity harmonic, J_4 , constrains the dynamics to the outermost 0.15 per cent of the total mass of Uranus and the outermost 0.2 per cent of the total mass of Neptune. This provides a stronger limit to the depth of the dynamical atmosphere than previously suggested⁶, and shows that the dynamics are confined to a thin weather layer no more than about 1,000 kilometres deep on both planets.

Measurable perturbations to the gravity fields of Uranus and Neptune can result from mass anomalies due to two sources—the rapid rotation of these planets, which distorts the planets into a non-spherical (oblate) shape, and density perturbations, which result from fast atmospheric winds^{6–10} organized on both planets into a broad zone of westward flow near the equator and eastward flow at high latitudes (Fig. 1). The gravity field can be decomposed into spherical gravity harmonics, (J_n), which are defined as a weighted integral over the planets' density distribution, $J_n = -(Ma^n)^{-1} \int P_n \rho r^n d^3r$, where P_n is the n th Legendre polynomial, M is the planetary mass, a is the mean planetary radius, ρ is the local density and r is the local radius¹¹. On planets with internal dynamics (winds), the density is perturbed by the flow so that the total density in J_n can be written as $\rho = \rho_{\text{static}} + \rho'$, where the density ρ_{static} is the hydrostatic density, and ρ' are the density fluctuations arising from internal dynamics. The gravity harmonics, can be then similarly decomposed into two parts $J_n = J_n^{\text{static}} + \Delta J_n^{\text{dyn}}$, where the first part (J_n^{static}) is due to the oblateness and radial density distribution of the planet and the second part (ΔJ_n^{dyn}) is due to the dynamical perturbations arising from winds^{8,10}.

In order to place an upper bound on the depth of the atmospheric circulation on Uranus and Neptune, we determine the difference between the observed J_4 and J_4^{static} resulting from wind-free models set to match all other observational constraints besides J_4 . Any difference in these quantities places constraints on the meteorological contribution to J_4 . Of course, the observed J_4 has uncertainties; moreover, there exists a full family of interior models ρ_{static} with different J_4^{static} values. Therefore, this difference can take on a wide range of possible values. Here we determine the maximum possible difference, which then determines the maximum possible contribution that dynamics can make to J_4 .

To determine the widest possible range of J_4^{static} values, we compute an ensemble of interior density profiles, $\rho_{\text{static}}(r)$, for each planet, calculated using the theory of figures¹², to third order, constrained to match the total mass, J_2 , mean radius, and the atmospheric density

and its derivative at a pressure level of 1 bar (see Supplementary Information). The resulting range of J_4^{static} values lie between -32.5×10^{-6} and -30.5×10^{-6} for Uranus, and between -34.5×10^{-6} and -31×10^{-6} for Neptune (Fig. 2 and Supplementary Fig. 1). Thus, even without constraining the interior models to the observed J_4 (for example, ref. 13), knowledge of the observed J_2 is enough to limit the possible solutions to be close to the observed J_4 values of $(-30.44 \pm 1.02) \times 10^{-6}$ and $(-33.40 \pm 2.90) \times 10^{-6}$ for Uranus and Neptune, respectively^{14,15}. Results from models using more sophisticated equations of state^{6,16–19} are within the same range for both planets. On the basis of the J_4^{static} values from the ensemble of interior models (Fig. 2 and Supplementary Fig. 1), and the observed values of J_4 (J_4^{observed}) including their uncertainties^{14,15}, and assuming $\Delta J_4^{\text{dyn}} = J_4^{\text{observed}} - J_4^{\text{static}}$, we find that ΔJ_4^{dyn} must be within the range $-1 \times 10^{-6} < \Delta J_4^{\text{dyn}} < 3 \times 10^{-6}$ for Uranus, and $-5 \times 10^{-6} < \Delta J_4^{\text{dyn}} < 4 \times 10^{-6}$ for Neptune.

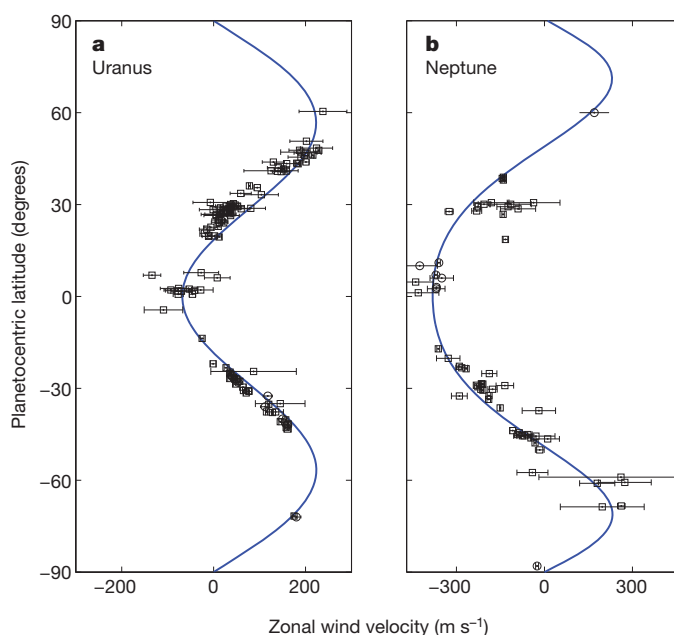


Figure 1 | Observed cloud-level zonally averaged zonal winds on Uranus and Neptune. **a**, Observations of Uranus from Voyager 2 (circles) and from HST measurements (squares)^{27,28}. The solid line is an empirical fit to the data²⁸. **b**, Observations of Neptune from Voyager 2 (circles)²⁹ and from HST measurements (squares)³⁰. The solid line is an empirical fit to the data²⁹, constrained to zero at the poles. The cloud-level atmospheric circulations on Uranus and Neptune have a generally similar structure, despite the differences in solar insolation (Uranus has an obliquity of 98° , whereas that of Neptune is 29°), and internal heating (Neptune's internal/solar heating ratio is roughly 1.6, whereas that of Uranus is only 0.06). Error bars represent cloud tracking and navigational errors^{27–30}.

¹Department for Environmental Sciences and Center for Planetary Science, Weizmann Institute of Science, Rehovot 76100, Israel. ²Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA. ³Department of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA. ⁴Department of Geophysics and Planetary Science, Tel Aviv University, Tel Aviv 69978, Israel.

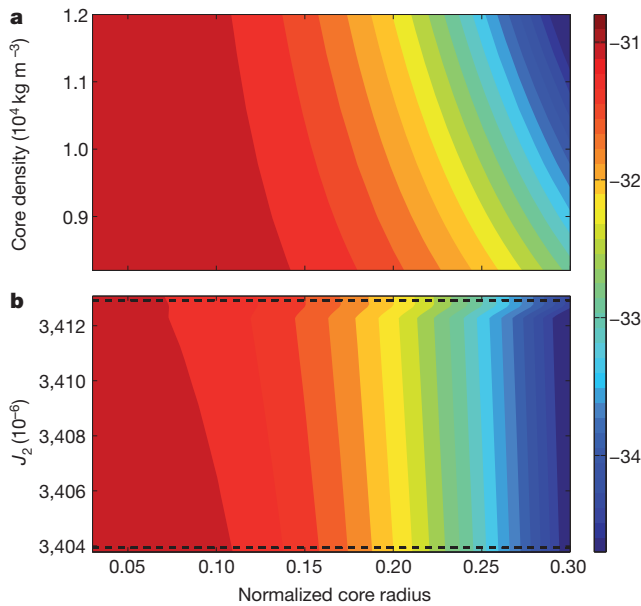


Figure 2 | J_4^{static} over a wide range of interior models for Neptune.

a. J_4^{static} as a function of normalized core radius and core density, with J_2 held constant at the mean observed value¹⁵ of $J_2 = 3,408.4 \times 10^{-6}$. **b.** J_4^{static} as a function of normalized core radius and J_2 , with the core density set to $1.01 \times 10^4 \text{ kg m}^{-3}$, and $J_2 = (3,408.4 \pm 4.5) \times 10^{-6}$ varying between the observed uncertainties (dashed lines). We specifically do not constrain the solution to J_4 , because we are interested in the possible range of J_4^{static} given the other constraints. We allow the constant-density core to extend up to 30% of the planet's radius (Fig. 3), its density to be up to $1.2 \times 10^4 \text{ kg m}^{-3}$ (refs 16, 17, 19) and J_2 to vary within the observed error estimates (see further details in Supplementary Information). A similar figure for Uranus appears as Supplementary Fig. 1.

Understanding the dynamical implications of these inferred ΔJ_4^{dyn} limit values requires knowledge of the zonal velocity structure. Because the planets are rapidly rotating, and Coriolis accelerations are dominant over inertial accelerations (small Rossby number), surfaces of constant angular momentum per unit mass will be nearly parallel to the axis of rotation^{20,21}. To leading order, this results in no interior flow crossing surfaces of constant angular momentum^{3,20,21}, and thus the fluid motion can be only along cylinders parallel to the spin axis, although the zonal wind velocity can decay towards the high-pressure interior^{20,22}. We therefore assume that the zonal wind field has the general form $u(r, \theta) = u_0 \exp[(r - a)/H]$ where $u_0(r, \theta)$ is the observed northern and southern hemisphere average cloud level zonal wind (Fig. 1) extended constantly along the direction of the axis of rotation (θ is latitude), and H is an e-folding decay depth of the cloud-level winds representing the possible shear of the winds^{8,10}. H is a free parameter, and varying it systematically allows exploration of the dependence of the gravity harmonics on the vertical extent of the winds. Thus, when $H \gg a$ the zonal wind is nearly constant along the direction of the axis of rotation, and as H is decreased the zonal velocity decreases more rapidly with depth^{8,10}. Because the dynamics are in the regime of small Rossby numbers, the flow to leading order is in geostrophic balance²³, and therefore the thermal wind balance must hold so that

$$(2\Omega \cdot \nabla)[\rho_{\text{static}} \mathbf{u}] = \nabla \rho' \times \mathbf{g}_0, \quad (1)$$

where Ω is the planetary rotation rate, $\mathbf{u}(\mathbf{r})$ is the full three-dimensional velocity and $\mathbf{g}_0(\mathbf{r})$ is the mean gravity vector^{20,23}. Here the thermal wind balance is written in a general form without making any assumptions about the depth of the circulation²⁰. Because the dynamics are a perturbation to the mean hydrostatic state, and the planets' deviation from spherical geometry is small (the equatorial radius is larger than the polar radius by 2.3% and 1.7% for Uranus and Neptune, respectively), we calculate the dynamical contribution to the gravity harmonics

(ΔJ_n^{dyn}) in spherical geometry. Thus, given the hydrostatic density $\rho_{\text{static}}(r)$ from interior models (for example, Fig. 3), the mean gravity \mathbf{g}_0 (which is calculated by integrating ρ_{static} radially) and the zonal velocity $u(r, \theta)$, the dynamical perturbation density $\rho'(r, \theta)$ can be calculated from the zonal component of equation (1), and will depend on the decay parameter H and an integration constant $\rho'_0(r)$. This integration constant has no contribution to ΔJ_n^{dyn} under spherical geometry, and has a negligible contribution in an oblate spheroid because $\rho'_0 \ll \rho_{\text{static}}$ (see Supplementary Information).

Comparing the allowable range of ΔJ_4^{dyn} inferred from $J_4^{\text{observed}} - J_4^{\text{static}}$ (dashed red lines in Fig. 4) to ΔJ_4^{dyn} calculated by the dynamical model with different wind depths allows placing an upper limit on the depth of the zonal winds. These values are calculated by systematically varying the decay depth H between 10 km and 10^5 km (thus from very shallow winds, to winds nearly penetrating the depth of the planet), calculating the resulting density perturbations (equation (1)) and then calculating ΔJ_4^{dyn} by integration over the spherical domain. We repeated this analysis for all models in our ensemble of interior models for $\rho_{\text{static}}(r)$, as well as for interior models of $\rho_{\text{static}}(r)$ inferred from more complex equations of state^{6,16,19}. All solutions lie between the blue curves in Fig. 4 for each of the planets.

Therefore, the largest possible depth of the flow where $\Delta J_4^{\text{dyn}} = 4 \times 10^{-6}$ for Neptune and $\Delta J_4^{\text{dyn}} = 3 \times 10^{-6}$ for Uranus occurs for shallow depths of roughly $H = 1,100$ km for both planets (Fig. 4). This means that the depth of the circulation on Neptune cannot exceed a pressure level of roughly 4,000 bar, which corresponds to the uppermost 0.2% of the total mass of the planet. Owing to the weaker winds on Uranus, the effect of the dynamics on ΔJ_4^{dyn} is smaller; however, because the planet is less massive the upper limit on the core size is lower (Supplementary Fig. 1), and therefore the maximum possible depth in Fig. 4 is similar to that of Neptune and corresponds to $\sim 2,000$ bar (which is roughly the outermost 0.15% of the planetary mass).

Previous studies using potential theory⁶ have shown that full differential rotation along cylinders is impossible for Neptune because the resulting J_4 will be large and positive ($\sim 10^{-4}$), whereas the observed value is negative. In the limit of deep winds (large H), our results for ΔJ_4^{dyn} match those results (Fig. 4). Here however, by using a continuous range of zonal velocity decay depths, and using today's better known observed

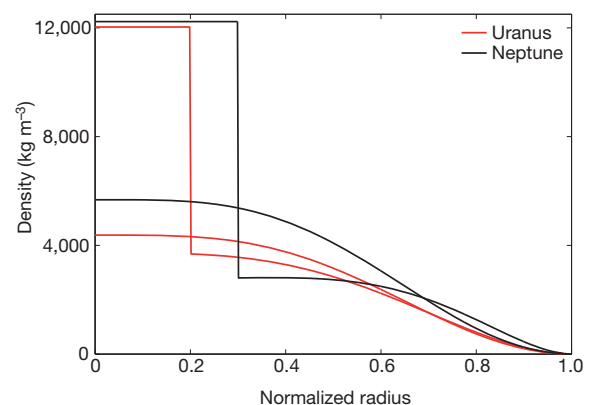


Figure 3 | Radial density profiles for two different interior models of

Uranus and Neptune. Interior profiles shown are the extreme cases of $\rho_{\text{static}}(r)$ from our suite of interior models. For each planet, we show here one model that has a constant core density of $\sim 1.2 \times 10^4 \text{ kg m}^{-3}$ reaching 30% of the planet's radius on Neptune (black) and 20% of the planet's radius on Uranus (red), and another model that does not have a constant density core. We used a suite of more than 3,000 profiles for Neptune and more than 1,500 profiles for Uranus, which are between these two extreme cases. All cases are constrained to match the planets' mass, J_2 , mean radius, and the atmospheric density and its derivative at 1 bar, but are not constrained to the observed J_4 (see Supplementary Information). Density profiles based on three-layer models^{6,16,19} were also used.

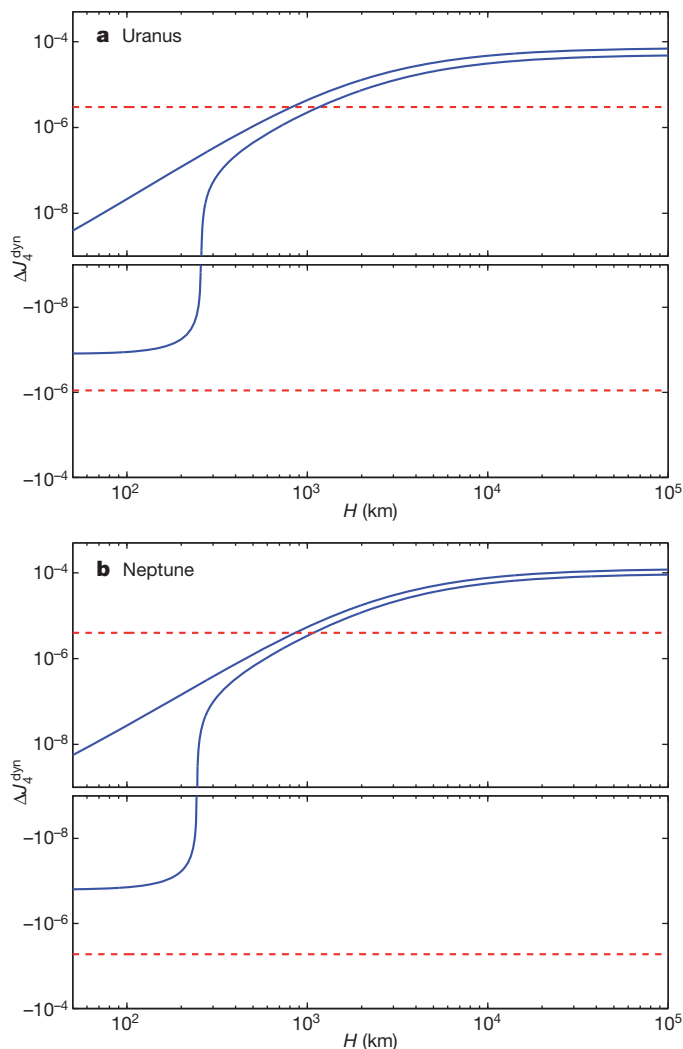


Figure 4 | ΔJ_4^{dyn} as function of the decay height H for Uranus and Neptune. **a**, Uranus; **b**, Neptune. All possible solutions for the range of interior models explored in this study are between the two blue lines for each planet. The dashed lines are the maximum and minimum possible values for ΔJ_4^{dyn} calculated as the difference between the observed J_4 and J_4^{static} obtained from the interior models (Fig. 2 and Supplementary Fig. 1). Only solutions within the two dashed red lines are possible solutions for the dynamical contribution to J_4 , and therefore H must be limited to less than $\sim 1,100$ km for both Uranus and Neptune. On Uranus, this depth corresponds to a pressure of roughly 2,000 bar or the outermost 0.15% of the mass. For Neptune, this is equivalent to a pressure of roughly 4,000 bar or the outermost 0.2% of the mass. For lower values of H (not shown), all ΔJ_4^{dyn} values converge to zero. For each planet, the bottom half of the plot is the negative of the log-scale to reflect the negative numbers on a log-scale.

values of J_4 (refs 14, 15), we provide much stronger constraints on the depth of the flow: we constrain the dynamics to the top few thousand bars, instead of the top few hundred thousand bars⁶.

The confinement of the strong jets on Uranus and Neptune to a shallow weather layer implies that the dynamics controlling zonal jets are likely to come from shallow processes, rather than from deep columnar structures that penetrate through the planet. Nevertheless, internal heat may be significant in driving these jets, particularly on Neptune where the internal heat flux is 1.6 times stronger than the heating from the Sun²⁴. Decay of the fast surface winds to small values within a shallow layer requires large horizontal density contrasts on isobars in the deep atmosphere. These could plausibly be provided by latent heating due to condensation of water at pressures of ~ 300 bar (ref. 2).

It is important to note that our results are an upper limit to the depth of the dynamics, owing to the conservative approach taken here of

using the widest range of reasonable interior models. These upper limits also hold when repeating the analysis with state-of-the-art three-layer interior structure models consisting of large ice/rock-rich cores¹⁹. Further understanding of the interior structures will probably narrow the range of possible J_4^{static} values, and thus confine the dynamics to be even shallower, although the precise constraint can depend on the details of the equation of state and density structure assumed. It is possible to imagine more complex dynamical scenarios (for example, where the depth of the winds varies in latitude). ΔJ_4^{dyn} is sensitive to the long-wavelength component of such variation, and, for this component, our results are robust and not model-dependent. In particular, the latitudinally averaged H must be smaller than the limits described in Fig. 4. We find these results to be robust to within a few per cent even when considering uncertainty in the exact rotation period^{25,26}. Although current knowledge of the gravity fields of Jupiter and Saturn is not sufficiently precise for a similar analysis, expected observations from the low-flying Juno and Cassini orbiters will enable tighter constraints on their low-order gravity fields, and hence on the depth of their dynamics^{8,10}.

Received 4 October 2012; accepted 22 March 2013.

- Read, P. L. Clearer circulation on Uranus. *Nature* **325**, 197–198 (1987).
- Lian, Y. & Showman, A. P. Generation of equatorial jets by large-scale latent heating on the giant planets. *Icarus* **207**, 373–393 (2010).
- Liu, J. & Schneider, T. Mechanisms of jet formation on the giant planets. *J. Atmos. Sci.* **67**, 3652–3672 (2010).
- Suomi, V. E., Limaye, S. S. & Johnson, D. R. High winds of Neptune — a possible mechanism. *Science* **251**, 929–932 (1991).
- Aurnou, J., Heimpel, M. & Wicht, J. The effects of vigorous mixing in a convective model of zonal flow on the ice giants. *Icarus* **190**, 110–126 (2007).
- Hubbard, W. B. *et al.* Interior structure of Neptune — comparison with Uranus. *Science* **253**, 648–651 (1991).
- Hubbard, W. B. Gravitational signature of Jupiter's deep zonal flows. *Icarus* **137**, 357–359 (1999).
- Kaspi, Y., Hubbard, W. B., Showman, A. P. & Flierl, G. R. Gravitational signature of Jupiter's internal dynamics. *Geophys. Res. Lett.* **37**, L01204 (2010).
- Kong, D., Zhang, K. & Schubert, G. On the variation of zonal gravity coefficients of a giant planet caused by its deep zonal flows. *Astrophys. J.* **748**, 143 (2012).
- Kaspi, Y. Inferring the depth of the zonal jets on Jupiter and Saturn from odd gravity harmonics. *Geophys. Res. Lett.* **40**, 676–680 (2013).
- Hubbard, W. B. *Planetary Interiors* (Van Nostrand Reinhold, 1984).
- Zharkov, V. N. & Trubitsyn, V. P. *Physics of Planetary Interiors* (Pachart Publishing House, 1978).
- Helled, R., Anderson, J. D., Podolak, M. & Schubert, G. Interior models of Uranus and Neptune. *Astrophys. J.* **726**, 15 (2011).
- Jacobson, R. A. The gravity field of the Uranian system and the orbits of the Uranian satellites and rings. *Bull. Am. Astron. Soc.* **39** (3), 453–453 (2007).
- Jacobson, R. A. The orbits of the Neptunian satellites and the orientation of the pole of Neptune. *Astrophys. J.* **137**, 4322–4329 (2009).
- Hubbard, W. B. & Marley, M. S. Optimized Jupiter, Saturn, and Uranus interior models. *Icarus* **78**, 102–118 (1989).
- Podolak, M., Weizman, A. & Marley, M. Comparative models of Uranus and Neptune. *Planet. Space Sci.* **43**, 1517–1522 (1995).
- Fortney, J. J. & Nettelmann, N. The interior structure, composition, and evolution of giant planets. *Space Sci. Rev.* **152**, 423–447 (2010).
- Nettelmann, N., Helled, R., Fortney, J. J. & Redmer, R. New indication for a dichotomy in the interior structure of Uranus and Neptune from the application of modified shape and rotation data. *Planet. Space Sci.* **77**, 143–151 (2013).
- Kaspi, Y., Flierl, G. R. & Showman, A. P. The deep wind structure of the giant planets: results from an anelastic general circulation model. *Icarus* **202**, 525–542 (2009).
- Schneider, T. & Liu, J. Formation of jets and equatorial superrotation on Jupiter. *J. Atmos. Sci.* **66**, 579–601 (2009).
- Liu, J., Goldreich, P. M. & Stevenson, D. J. Constraints on deep-seated zonal winds inside Jupiter and Saturn. *Icarus* **196**, 653–664 (2008).
- Pedlosky, J. *Geophysical Fluid Dynamics* (Springer, 1987).
- Pearl, J. C. & Conrath, B. J. The albedo, effective temperature, and energy balance of Neptune, as determined from Voyager data. *J. Geophys. Res.* **96**, 18921–18930 (1991).
- Helled, R., Anderson, J. D. & Schubert, G. Uranus and Neptune: shape and rotation. *Icarus* **210**, 446–454 (2010).
- Karkoschka, E. Neptune's rotational period suggested by the extraordinary stability of two features. *Icarus* **215**, 439–448 (2011).
- Hammel, H. B., de Pater, I., Gibbard, S., Lockwood, G. W. & Rages, K. Uranus in 2003: zonal winds, banded structure, and discrete features. *Icarus* **175**, 534–545 (2005).
- Sromovsky, L. A. & Fry, P. M. Dynamics of cloud features on Uranus. *Icarus* **179**, 459–484 (2005).

29. Sromovsky, L. A., Limaye, S. S. & Fry, P. M. Dynamics of Neptune's major cloud features. *Icarus* **105**, 110–141 (1993).
30. Sromovsky, L. A., Fry, P. M., Dowling, T. E., Baines, K. H. & Limaye, S. S. Neptune's atmospheric circulation and cloud morphology: changes revealed by 1998 HST imaging. *Icarus* **150**, 244–260 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements Y.K. and O.A. thank the Helen Kimmel Center for Planetary Science at the Weizmann Institute of Science for support. A.P.S. and W.B.H. acknowledge support by NASA.

Author Contributions Y.K. and A.P.S. initiated and designed the research. Y.K. performed the dynamical gravity harmonics calculations and wrote the paper. R.H. performed the static interior model calculations and their interpretation. All authors contributed to the discussion of the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.K. (yohai.kaspi@weizmann.ac.il).

An electrically pumped polariton laser

Christian Schneider^{1*}, Arash Rahimi-Iman^{1*}, Na Young Kim^{2,3}, Julian Fischer¹, Ivan G. Savenko^{4,5}, Matthias Amthor¹, Matthias Lerner¹, Adriana Wolf¹, Lukas Worschech¹, Vladimir D. Kulakovskii⁶, Ivan A. Shelykh^{4,5}, Martin Kamp¹, Stephan Reitzenstein^{1,7}, Alfred Forchel¹, Yoshihisa Yamamoto^{2,8} & Sven Höfling¹

Conventional semiconductor laser emission relies on stimulated emission of photons^{1,2}, which sets stringent requirements on the minimum amount of energy necessary for its operation^{3,4}. In comparison, exciton–polaritons in strongly coupled quantum well microcavities⁵ can undergo stimulated scattering that promises more energy-efficient generation of coherent light by ‘polariton lasers’^{3,6}. Polariton laser operation has been demonstrated in optically pumped semiconductor microcavities at temperatures up to room temperature^{7–12}, and such lasers can outperform their weak-coupling counterparts in that they have a lower threshold density^{12,13}. Even though polariton diodes have been realized^{14–16}, electrically pumped polariton laser operation, which is essential for practical applications, has not been achieved until now. Here we present an electrically pumped polariton laser based on a microcavity containing multiple quantum wells. To prove polariton laser emission unambiguously, we apply a magnetic field and probe the hybrid light–matter nature of the polaritons. Our results represent an important step towards the practical implementation of polaritonic light sources and electrically injected condensates, and can be extended to room-temperature operation using wide-bandgap materials.

Quantum well microcavity exciton–polaritons are composite bosons consisting partly of light and partly of matter. They originate from the strong-coupling regime between excitons and photons, and therefore mix the properties of these excitations^{5,6}. Exciton–polaritons can undergo dynamical condensation above a critical particle density^{7–12}. In the underlying nonlinear process, a macroscopic number of polaritons is accumulated in the single-particle ground state of the lower-energy dispersion branch of polaritons (‘lower polaritons’) by stimulated polariton scattering. Exciton–polaritons in such a condensate decay by the leakage of photons from the microcavity, which produces monochromatic and coherent light.

In this work, we demonstrate nonlinear emission by an electrically driven microcavity device operating in the strong-coupling regime. Because such a device has similarities with a weakly coupled microcavity laser^{17,18}, care must be taken to demonstrate polariton laser operation unambiguously. Therefore, we collected a range of evidence showing that the strong-coupling regime is preserved in our device across its polariton laser threshold, as follows. We observed three different regimes in the energy–momentum dispersion characteristics, which can be attributed to the regimes of incoherent polariton emission, polariton laser operation and cavity mediated laser operation, respectively. The corresponding two transitions are manifest in the input–output characteristics as nonlinearities. This behaviour is especially pronounced in an applied magnetic field. The exciton density at the polariton laser threshold is significantly lower than the carrier density required by the Bernard–Duraffourg condition⁴ for conventional lasers. Finally, we go a step beyond these standard characterization criteria and apply a magnetic field in Faraday configuration to probe the magnetic properties of the polariton system with crucial

implications. Apart from a clear enhancement of the nonlinearity related to the polariton laser threshold¹⁹, we observe a significant Zeeman splitting of the fundamental mode in the polariton laser regime. This reflects the hybrid character of the system and shows that strong coupling is preserved across the first—polariton-laser-related—threshold.

Our exciton–polariton laser diode (Fig. 1a) is based on a microcavity structure with 23 GaAs/AlAs mirror pairs in the top distributed Bragg reflector (DBR), 27 GaAs/AlAs mirror pairs in the bottom DBR and four integrated InGaAs quantum wells in an intrinsic GaAs one-wavelength-thick cavity (Methods Summary). We characterized the wafer via reflection measurements by using the spatial variation of the cavity resonance (Fig. 1b), extracting a vacuum Rabi splitting of $2\hbar\Omega = 5.5 \pm 0.2$ meV (\hbar , Planck’s constant divided by 2π) in the presence of the built-in potential from the p–i–n junction (which comprises p- and n-type regions separated by an undoped region). To extract the quality factor (Q-factor) of the microresonator, we fabricated micropillar cavities with diameters of 2 μm on a far-red-detuned area of the wafer (Supplementary Information). Microphotoluminescence spectra of these structures yield a value of $Q = 6,320 \pm 60$ (Fig. 1c). Current confinement in our injection scheme is provided by fabricating electrically contacted circular pillars of diameter 20 μm (Fig. 1a and Methods).

We investigate the polaritonic electroluminescence emission of our device by studying the current-density-dependent energy–momentum dispersions at cavity–exciton detunings of $\Delta = E_{\text{C0}} - E_{\text{X0}} = -4.8 \pm 0.3$ meV for zero magnetic field (0 T) and -6.5 ± 0.3 meV for 5 T. These detunings correspond to excitonic fractions at momentum $k = 0$ of $|X|^2 \approx 17\%$ for 0 T and 13% for 5 T (Fig. 2 and Supplementary Information). Different excitation densities of ~ 0.5 –0.9, 1.4 and 2.9–3.1 times the assigned polariton laser threshold current densities, $j_{\text{th1}} = 82 \pm 5$ A cm^{−2} at 0 T and $j_{\text{th1}} = 77 \pm 2$ A cm^{−2} at 5 T, are depicted. Spectra recorded at 0 T are plotted in Fig. 2a–c, and the corresponding dispersions of the same device recorded at 5 T are depicted in Fig. 2d–f. For all applied pump currents, the higher-energy dispersion branch of polaritons (‘upper polaritons’) is thermalized owing to the high Q-factor of our device^{20,21}. Hence, the polaritonic character of the emission has to be assessed by analysing the characteristics of the lower-polariton branch. In Fig. 2a, d, the theoretical dispersion of the uncoupled cavity mode (C, dashed yellow) with $E_{\text{C0}} = 1.4166$ eV is projected together with the corresponding lower-polariton dispersion (dashed black). In the region of the dispersion around the lower-polariton ground state with high photonic content ($\sim 83\%$ at 0 T and $\sim 87\%$ at 5 T) and low effective masses, quantization of the lower-polariton mode is observed with a mode separation of the order of the emission mode’s linewidth, which amounts to 0.24 \pm 0.03 meV at low pump densities.

The spectral signatures in Fig. 2a, b show that polaritons are efficiently scattered into the lowest-energy state in the absence of a magnetic field.

¹Technische Physik und Wilhelm-Conrad-Röntgen-Research Center for Complex Material Systems, Universität Würzburg, D-97074 Würzburg, Am Hubland, Germany. ²E. L. Ginzton Laboratory, Stanford University, Stanford, California 94305, USA. ³Institute of Industrial Science, University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan. ⁴Science Institute, University of Iceland, Dunhaga 3, IS-107 Reykjavik, Iceland. ⁵Division of Physics and Applied Physics, Nanyang Technological University, 637371, Singapore. ⁶Institute of Solid State Physics, Russian Academy of Science, Chernogolovka 142432, Russia. ⁷Institut für Festkörperphysik, Technische Universität Berlin, Hardenbergstraße 36, D-10623 Berlin, Germany. ⁸National Institute of Informatics, Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan.

*These authors contributed equally to this work.

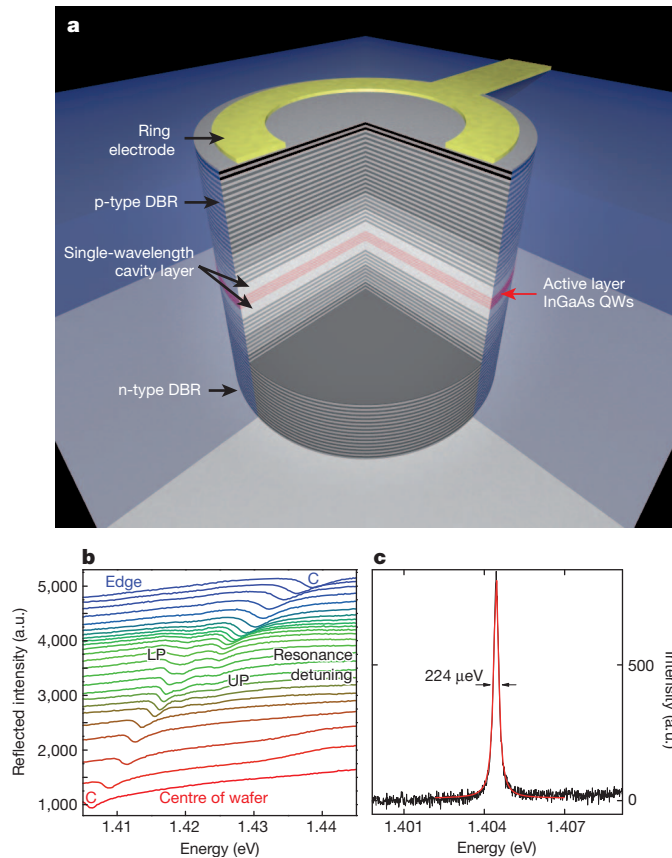


Figure 1 | Quantum well microcavity polariton diode and characteristics. **a**, Schematics of an electrically contacted 20-μm-diameter micropillar with four quantum wells (QWs) in the cavity, sandwiched by gradually doped distributed Bragg reflectors (DBRs). **b**, Waterfall plot of reflectivity spectra showing anticrossing on the detuning map. a.u., arbitrary units; LP, lower polariton; UP, upper polariton; C, photonic cavity mode. **c**, Microphotoluminescence spectrum of the fundamental mode of a highly photonic micropillar cavity with a diameter of 2 μm and $Q \approx 6,320$.

This behaviour changes qualitatively when an external magnetic field of 5 T is applied. For low-excitation conditions (Fig. 2d), the polariton scattering rate into the ground state is lower than that for 0 T, resulting in the occurrence of an emission bottleneck at $k \approx \pm 1.5 \mu\text{m}^{-1}$ and a reduction in intensity at $k = 0$. Remarkably, this bottleneck is overcome at higher pump currents (Fig. 2e), and the polariton ground state is very highly occupied. This behaviour is further reflected in the polariton occupation as a function of energy (Supplementary Information). The enhanced macroscopic ground-state occupation is one important characteristic of a polariton laser. At higher pump rates, of about $3j_{\text{th1}}$, a further notable change in the dispersion is revealed: the ground-state emission energy is blueshifted by 0.4 meV with respect to the polariton laser energy just above j_{th1} , due to a transition into the weak-coupling regime, and photonic resonances with resolution-limited linewidth dominate the spectrum (Fig. 2c, f). These features are usually associated with the photon laser regime^{12,13,22–24}. Recently, there have been indications that the quantum well microcavity system can undergo a transition into a photonic Bose–Einstein condensate in the weak-coupling regime^{18,24}, and theoretical work has also predicted a transition to a photon-dominated Bardeen–Cooper–Schrieffer regime of weakly correlated electron–hole pairs^{25,26}. A full understanding of the nature of this second transition hence requires further investigation.

To investigate these two transitions quantitatively, in Fig. 3a we plot the emitted photon flux at $k = 0 \pm 0.1 \mu\text{m}^{-1}$ as a function of the excitation current density. The device was again investigated with and without an applied magnetic field. At 5 T, two thresholds are indicated in the device's input–output characteristics, occurring at

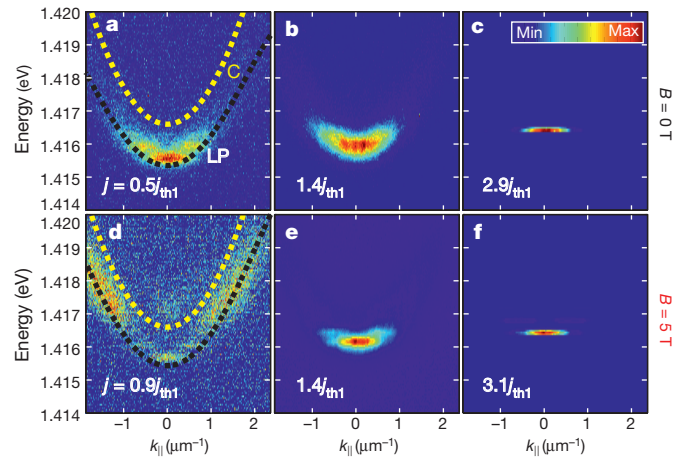


Figure 2 | Spectral emission features in various excitation regimes. **a–c**, Energy–momentum dispersions ($k_{||}$, in-plane wavevector) with false-colour intensity profiles (normalized to maximum intensity) corresponding to current densities, j , of ~ 44 , ~ 113 and $\sim 241 \text{ A cm}^{-2}$ at 0 T. Below threshold, the polaritonic system is characterized by a thermal distribution of particles (**a**). High ground-state population is observed at higher injection currents (**b**). At even higher pump rates, photonic lasing occurs (**c**). **d–f**, Corresponding dispersions at 5 T exhibiting similar, yet more-pronounced, characteristic transition features.

respective current densities of $77 \pm 2 \text{ A cm}^{-2}$ and $190 \pm 10 \text{ A cm}^{-2}$, which manifest themselves as nonlinearities in the log–log scale^{13,22–24}. In the absence of a magnetic field, the first nonlinearity can be extracted only from a slight but noticeable deviation from the extrapolated input–output curve at low pump current densities ($< 75 \text{ A cm}^{-2}$), and occurs at a current density of $82 \pm 5 \text{ A cm}^{-2}$. At 5 T, a drastic, nonlinear intensity increase is observed. We attribute this strong modification of the input–output characteristics by the magnetic field to an interplay of two effects: the increase in the exciton–photon detuning via the quantum well diamagnetic shift (Supplementary Information); and the enhancement of the excitonic oscillator strength by $\sim 15\%$ at 5 T (Supplementary Information), which causes a strong modification of the scattering processes in the lower-polariton branch. From the given current densities, we can estimate the exciton density per quantum well to be $(6 \pm 3) \times 10^9 \text{ cm}^{-2}$ (Supplementary Information) at the first transition at 5 T, which is lower than the exciton density in the transition regime between an exciton gas to an electron–hole plasma²⁷ (typically of the order of $(1–10) \times 10^{10} \text{ cm}^{-2}$ for comparable quantum wells). Conclusively, this transition is compatible with the strong-coupling regime. At the second transition, the current density, $j_{\text{th2}} \approx 190 \pm 10 \text{ A cm}^{-2}$, yields an exciton density of $\sim (1.5 \pm 0.7) \times 10^{10} \text{ cm}^{-2}$, for which the oscillator strength is already bleached—that is, has decreased—owing to exciton screening effects at high densities, and the system enters the weak coupling regime.

The excitation-dependent evolution of the emission energy and linewidth is shown in Fig. 3b for 0 T and Fig. 3c for 5 T. The open circles in the contour maps represent the peak emission energies of the ground-state mode. Starting from the lower-polariton energy for both 0 T and 5 T in the linear regime, an overall blueshift of the emission ground state by $\sim 0.8 \text{ meV}$ is measured up to the mode pinning at an energy $\sim 1.4164 \text{ eV}$ above the second threshold. The blueshift in the nonlinear polaritonic regime is partly explained by the population-dependent interaction energy in the macroscopically occupied ground state^{8,9}. Furthermore, the reduction in the exciton oscillator strength around the Mott density blueshifts the emission towards the uncoupled-photon energy, and a magnetic-field-dependent diamagnetic shift of $\sim 0.16 \text{ meV}$ develops for 5 T (see below).

As a result of the dominant photonic component in the red-detuned polariton system ($\mathcal{A} \approx -|2\hbar\Omega|$), we observe a rather narrow linewidth

of 0.24 ± 0.03 meV below threshold. After the transition at j_{th1} , the linewidth remains at a constant level of 0.30 ± 0.03 meV, representing the linewidth of exciton–polaritons with little matter content and, thus, little particle–particle interaction. Before the typical broadening related to polariton–polariton interaction at higher excitation is observable⁹, continuous bleaching of strong coupling takes place, the photonic content becomes increasingly dominant and the system crosses the second threshold to reach a resolution-limited emission linewidth of ~ 0.05 meV.

So far in this report, the magnetic field has served as a tool to enhance the polariton scattering efficiency in our microcavity device¹⁹. However, we also use the magnetic field as a probe to characterize unambiguously the emission character of the hybrid light–matter quasiparticles: photons emitted from a polaritonic system exhibit a specific magnetic-field-dependent Zeeman splitting both below²⁸ and above the polariton laser threshold²⁹, whereas cavity emission in the weak coupling regime is unaffected by a magnetic field. Therefore, the system's response to a magnetic field can be exploited to characterize the nature of the emission.

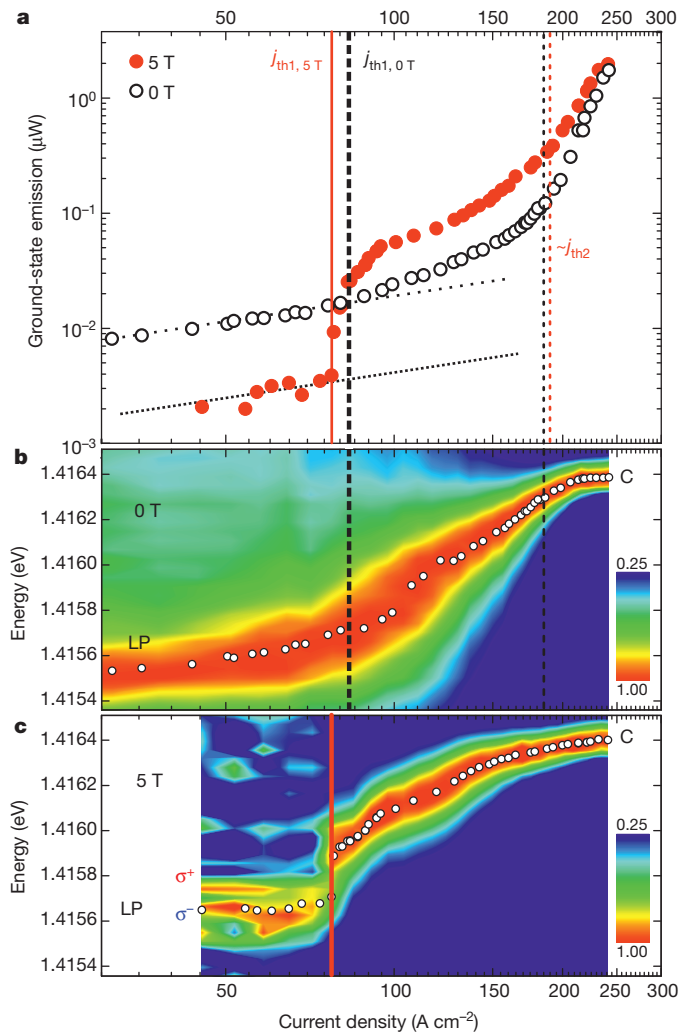


Figure 3 | Current-density dependency of the polariton diode emission. **a**, Input–output curves for 0 and 5 T on a double-logarithmic scale. Vertical lines mark two distinct thresholds: j_{th1} , which lies at 77 A cm^{-2} for 5 T (solid red), and j_{th2} , which lies at 190 A cm^{-2} for 5 T (dotted red). The corresponding thresholds for 0 T are marked by the vertical black lines. Oblique dotted lines indicate the linear increase in intensity in the low-density regime. **b**, **c**, Contour plots of the evolution of the device's emission linewidth as a function of the excitation current at 0 T (**b**) and 5 T (**c**). The spectra were normalized to maximum intensity for clarity. Note that the energy and colour scales are linear.

To characterize the three different operating regimes of our device, we present polarization-selective and magnetic-field-dependent emission spectra in Fig. 4. Figure 4a shows the ground-state signal of the lower polariton below the polariton lasing threshold, Fig. 4b shows this signal in the assigned polariton laser regime and Fig. 4c shows it in the highly photonic regime. In Fig. 4a, the energetic difference between the circularly σ^+ - and σ^- -polarized light yields the Zeeman splitting, which monotonically increases with increasing magnetic field. The

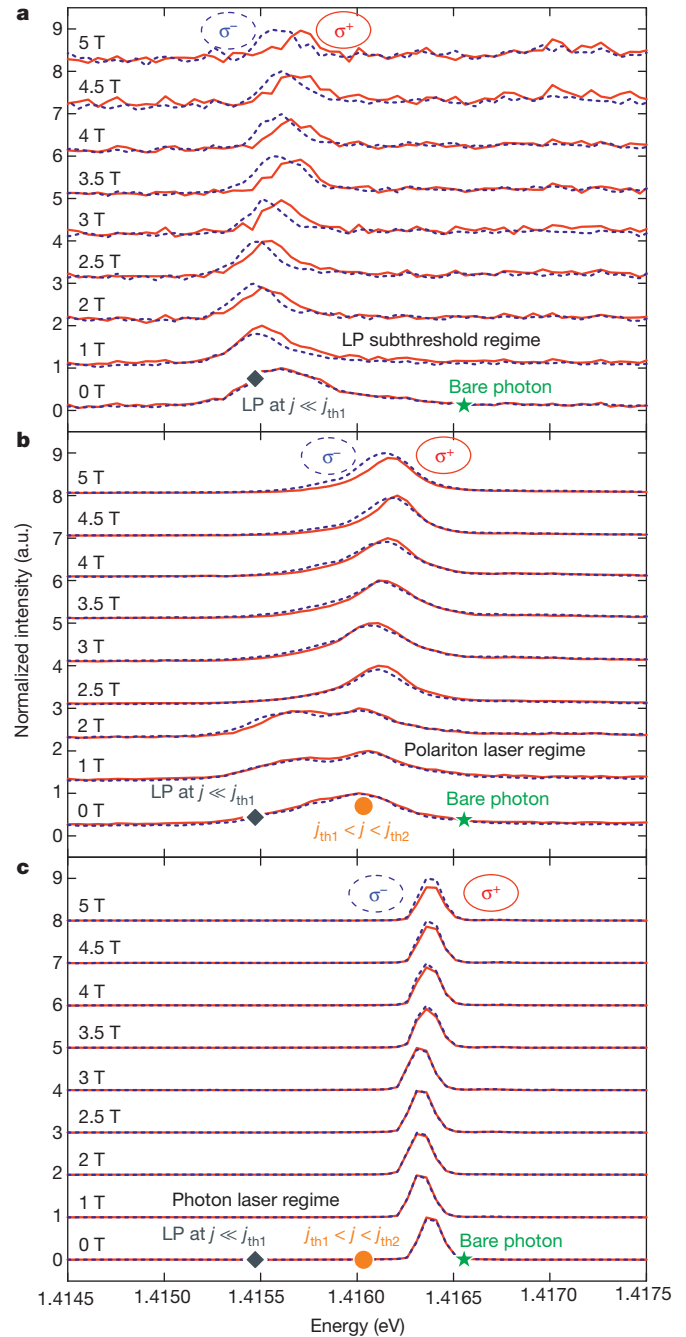


Figure 4 | Magnetic-field-dependent circular polarization spectra. **a**, Circularly σ^+ - and σ^- -polarized spectra of the lower-polariton emission at various magnetic fields below the polariton lasing threshold at $\sim 65 \text{ A cm}^{-2}$. The very low-pump-power lower-polariton ground-state emission at 0 T and the bare-cavity mode are marked by a diamond and a star, respectively. **b**, Polarized spectra representing emission in the polariton laser regime, recorded at a current density corresponding to 60% of the entire blueshift towards the bare-cavity mode (marked at 0 T by an orange circle). **c**, Device electroluminescence in the weak-coupling lasing regime.

device emitting in the linear regime at various magnetic fields was driven by an excitation current density of $\sim 65 \text{ A cm}^{-2}$. In addition to the Zeeman splitting, the modes experience a diamagnetic blueshift of $0.16 \pm 0.02 \text{ meV}$ (Supplementary Information). Above the polariton laser threshold, a less pronounced but visible mode splitting can be observed for various magnetic fields in the polarization-resolved spectra displayed in Fig. 4b. This indicates a polariton-interaction-dependent decrease in the Zeeman splitting in the polariton lasing regime, evidencing an excitonic component (see below). Figure 4c presents polarization-resolved, magnetic-field-dependent spectra, recorded above the second threshold, of the corresponding far-field spectra, which are depicted in Fig. 2c, f. As expected, no magnetic-field-dependent mode splitting can be observed.

The characteristic splitting of the polarized modes of these three regimes is depicted in Fig. 5a. Here the expected linear dependence of the Zeeman splitting as a function of the applied magnetic field below the polariton lasing threshold (Fig. 5a, blue squares) and the characteristics above the threshold (Fig. 5a, red circles) can be identified. In agreement with Fig. 4b, the splitting is less in the polariton lasing regime. As a remarkable feature, in this regime, quenching of the Zeeman splitting in accordance with theoretical considerations is obtained for fields up to 2 T (ref 30.), where compensation of the splitting by polariton–polariton interactions is predicted. Above the second transition in the weak-coupling regime of our device (Fig. 5a, green stars), the Zeeman splitting is completely bleached.

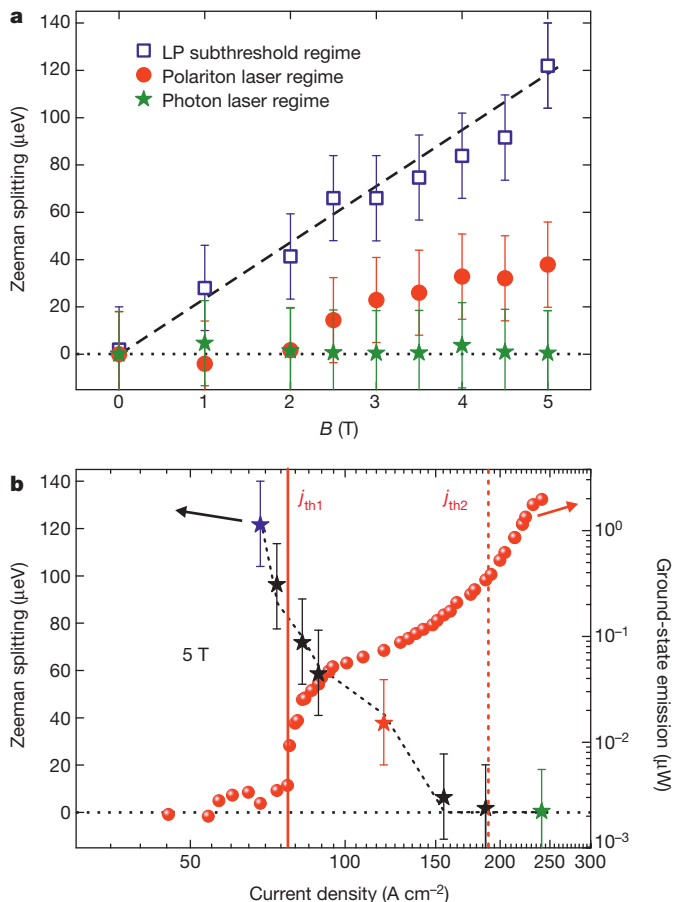


Figure 5 | Zeeman splitting of the polaritonic emission. **a**, Mode splitting as a function of magnetic field for different current densities in the three characteristic regimes. The displayed error bars of $\pm 20 \mu\text{eV}$ represent the estimated maximum errors in determining splittings from the underlying mode energies. **b**, Current-density-dependent mode splitting at 5 T. The input–output characteristic from Fig. 3a is plotted in the same graph to visualize the two transitions.

We additionally probe the excitation-current dependency of the Zeeman splitting by measuring the characteristic ground-state mode splitting at a magnetic field of 5 T at $k = 0 \pm 0.1 \mu\text{m}^{-1}$ as a function of the excitation current. This magnetic field is beyond the critical magnetic field for which a quenching of the Zeeman splitting in the polariton lasing regime is expected^{29,30}. Above this critical magnetic field, the Zeeman splitting of a polaritonic system is approximated by $\Delta E_{LP}(B, |X|^2) = E_{\sigma+} - E_{\sigma-} = |X|^2 g_X \mu_B (B - B_{cr})$ (ref 30.), where μ_B is the Bohr magneton, g_X is the g -factor of the exciton and $B_{cr} = (\alpha_1 - \alpha_2)n/|X|^2 g_X \mu_B$ is a critical value of the field proportional to the ground-state occupancy, n . In B_{cr} , $\alpha_1 = |X|^4 6E_{bind}a_B^2$ and $\alpha_2 \approx -0.1\alpha_1$ denote the Coulomb scattering parameters for polaritons with parallel and, respectively, antiparallel spins^{30,31}, and here E_{bind} and a_B denote the binding energy and the Bohr radius of the excitons, respectively. The formula describing the Zeeman splitting, which depends on the magnetic-field- and excitation-dependent polariton Hopfield coefficient and the spin interaction coefficients, yields an autocorrelation between the splitting and the polariton density.

Figure 5b shows experimental data on the Zeeman splitting plotted as a function of injection current, which can be modelled using this theoretical model together with the input–output characteristics of the device recorded at 5 T. Below the polariton laser threshold, we observe a Zeeman splitting of $\sim 120 \pm 20 \mu\text{eV}$ at 5 T, corresponding to $g_X \mu_B \approx 180 \mu\text{eV T}^{-1}$. This is in good agreement with values from a comparable single quantum well of similar design with $g_X \mu_B \approx 110 \mu\text{eV T}^{-1}$ (ref 28.), because this factor can vary for similar systems as a result of dependences on the effective quantum well thickness and composition³². Once the polariton laser threshold in the input–output characteristics is crossed, the Zeeman splitting at 5 T continuously decreases owing to the aforementioned density dependence of B_{cr} . The fact that the mode splitting at 5 T remains detectable up to the current at which the photon lasing threshold is crossed is unambiguous evidence of an excitonic component of our system, due to which the strong coupling is preserved across the first threshold attributed to polariton lasing.

We have observed clear evidence for polariton laser emission in an electrically driven quantum well microcavity device. In addition to standard characterization methods for polariton lasers, we have identified a clear-cut criterion to distinguish between a polariton-mediated laser and a cavity-mediated laser by characterizing the Zeeman splitting of the emission mode. We anticipate that this work can be extended to the realization of compact, electrically pumped Bose–Einstein condensates, which should be observable even up to room temperature in wide-bandgap material systems.

METHODS SUMMARY

Sample design and fabrication. The sample was grown by molecular beam epitaxy on (100)-oriented silicon-doped (n-doped) GaAs substrates. A stack of four 8-nm-thick $\text{In}_{0.15}\text{Ga}_{0.85}\text{As}$ quantum wells, separated by 6-nm-thick GaAs barriers, was integrated into an undoped, one-wavelength-thick (281-nm) GaAs cavity, which was centred between 23 GaAs (64 nm)/AlAs (71 nm) mirror pairs in the top DBR and 27 mirror pairs in the bottom DBR. The doping in the mirrors was symmetrically reduced from 3×10^{18} to $1 \times 10^{18} \text{ cm}^{-3}$ along both the n-type and the p-type sections towards the intrinsic cavity. In the topmost two p-type mirror pairs, the doping concentration was increased to $2 \times 10^{19} \text{ cm}^{-3}$. Delta-doped layers (sheet density, 10^{12} cm^{-2}) were included at every second interface to improve the electrical properties of the devices. The sample was patterned into circular pillars with diameters of $20 \mu\text{m}$ by electron cyclotron resonance reactive ion etching. After planarization using benzocyclobutene, ring-shaped p-contacts (Ti–Au) were evaporated at the upper facets of the pillars, and an additional semi-transparent metal film (Ti (3 nm)/Au (7 nm)) was deposited on the p-type side. The close stack of InGaAs quantum wells guarantees homogeneous carrier injection into each quantum well. The backside n-type contact consists of a AuGe–Ni–Au alloy.

Experimental set-up. A magneto-optical set-up was constructed in which both real-space and momentum-space resolved spectroscopy and imaging are accessible. Electroluminescence data was collected through a 0.4 NA microscope objective and directed into an imaging spectrometer using a set of lenses, to project the

proper projection plane onto the monochromator's entrance slit. The system's momentum-space resolution was $\sim 0.05 \mu\text{m}^{-1}$ ($\sim 0.5^\circ$) and the spectral resolution was $\sim 0.05 \text{ meV}$. We used a nitrogen-cooled Si-CCD as detector. The momentum-space coverage of this configuration amounts to $\pm 2.2 \mu\text{m}^{-1}$. A superconducting magnetocryostat was used to apply Faraday magnetic fields.

Received 7 November 2012; accepted 21 February 2013.

1. Coldren, L. A. & Corzine, S. W. *Diode Lasers and Photonic Integrated Circuits* (Wiley, 1995).
2. Einstein, A. Strahlungs-emission und -absorption nach der Quantentheorie. *Verh. Deutsch. Phys. Gesell.* **18**, 318–323 (1916).
3. Imamoğlu, A., Ram, R. J., Pau, S. & Yamamoto, Y. Nonequilibrium condensates and lasers without inversion: exciton-polariton lasers. *Phys. Rev. A* **53**, 4250–4253 (1996).
4. Bernard, M. G. A. & Durauffourg, G. Laser conditions in semiconductors. *Phys. Status Solidi B* **1**, 699–703 (1961).
5. Weisbuch, C., Nishioka, M., Ishikawa, A. & Arakawa, Y. Observation of the coupled exciton-photon mode splitting in a semiconductor quantum microcavity. *Phys. Rev. Lett.* **69**, 3314–3317 (1992).
6. Kavokin, A. & Malpuech, G. *Cavity Polaritons* (Elsevier, 2003).
7. Deng, H. *et al.* Condensation of semiconductor microcavity exciton polaritons. *Science* **298**, 199–202 (2002).
8. Kasprzak, J. *et al.* Bose–Einstein condensation of exciton polaritons. *Nature* **443**, 409–414 (2006).
9. Balili, R. *et al.* Bose–Einstein condensation of microcavity polaritons in a trap. *Science* **316**, 1007–1010 (2007).
10. Christopoulos, S. *et al.* Room-temperature polariton lasing in semiconductor microcavities. *Phys. Rev. Lett.* **98**, 126405 (2007).
11. Sun, L. *et al.* Room temperature one-dimensional polariton condensate in a ZnO microwire. Preprint at <http://arxiv.org/abs/1007.4686> (2010).
12. Deng, H., Weihs, G., Snoke, D., Bloch, J. & Yamamoto, Y. Polariton lasing vs. photon lasing in a semiconductor microcavity. *Proc. Natl Acad. Sci. USA* **100**, 15318–15323 (2003).
13. Tsotsis, J. *et al.* Lasing threshold doubling at the crossover from strong to weak coupling regime in GaAs microcavity. *N. J. Phys.* **14**, 023060 (2012).
14. Tsiatzos, S. I. *et al.* A GaAs polariton light-emitting diode operating near room temperature. *Nature* **453**, 372–375 (2008).
15. Bajoni, D. *et al.* Polariton light-emitting diode in a GaAs-based microcavity. *Phys. Rev. B* **77**, 113303 (2008).
16. Khalifa, A. A., Love, A. P. D., Krizhanovskii, D. N., Skolnick, M. S. & Roberts, J. S. Electroluminescence emission from polariton states in GaAs-based semiconductor microcavities. *Appl. Phys. Lett.* **92**, 061107 (2008).
17. Bajoni, D., Senellart, P., Lemaître, A. & Bloch, J. Photon lasing in GaAs microcavity: similarities with a polariton condensate. *Phys. Rev. B* **76**, 201305 (2007).
18. Ohadi, H. *et al.* Spontaneous symmetry breaking in a polariton and photon laser. *Phys. Rev. Lett.* **109**, 016404 (2012).
19. Kulakovskii, V. D. *et al.* Magnetic field control of polarized polariton condensates in rectangular microcavity pillars. *Phys. Rev. B* **85**, 155322 (2012).
20. Wertz, E. *et al.* Spontaneous formation of a polariton condensate in a planar GaAs microcavity. *Appl. Phys. Lett.* **95**, 051108 (2009).
21. Kulakovskii, V. D. *et al.* Bose–Einstein condensation of exciton polaritons in high-Q planar microcavities with GaAs quantum wells. *JETP Lett.* **92**, 595–599 (2010).
22. Nelsen, B., Balili, R., Snoke, D. W., Pfeiffer, L. & West, K. Lasing and polariton condensation: two distinct transitions in GaAs microcavities with stress traps. *J. Appl. Phys.* **105**, 122414 (2009).
23. Dang, L. S., Heger, D., Andre, R., Boeuf, F. & Romestain, R. Stimulated emission of polariton luminescence in semiconductor microcavity. *Phys. Rev. Lett.* **81**, 3920–3923 (1998).
24. Kammann, E., Ohadi, H., Maragkou, M., Kavokin, A. V. & Lagoudakis, P. G. Crossover from photon to exciton-polariton lasing. *N. J. Phys.* **14**, 105003 (2012).
25. Keeling, J., Eastham, P. R., Szymanska, M. H. & Littlewood, P. B. BCS-BEC crossover in a system of microcavity polaritons. *Phys. Rev. B* **72**, 115320 (2005).
26. Byrnes, T., Horikiri, T., Ishida, N. & Yamamoto, Y. BCS wavefunction approach to the BEC-BCS crossover of exciton-polariton condensates. *Phys. Rev. Lett.* **105**, 186402 (2010).
27. Kappei, L., Szczytko, J., Morier-Genoud, F. & Deveaud, B. Direct observation of the Mott transition in an optically excited semiconductor quantum well. *Phys. Rev. Lett.* **94**, 147403 (2005).
28. Rahimi-Iman, A. *et al.* Zeeman splitting and diamagnetic shift of spatially confined quantum-well exciton polaritons in an external magnetic field. *Phys. Rev. B* **84**, 165325 (2011).
29. Larionov, A. V. *et al.* Polarized nonequilibrium Bose–Einstein condensates of spinor exciton polaritons in a magnetic field. *Phys. Rev. Lett.* **105**, 256401 (2010).
30. Rubo, Y. G., Kavokin, A. V. & Shelykh, I. A. Suppression of superfluidity of exciton-polaritons by magnetic field. *Phys. Lett. A* **358**, 227–230 (2006).
31. Tassone, F. & Yamamoto, Y. Exciton-exciton scattering dynamics in a semiconductor microcavity and stimulated scattering into polaritons. *Phys. Rev. B* **59**, 10830–10842 (1999).
32. Kotlyar, R., Reinecke, T. L., Bayer, M. & Forchel, A. Zeeman spin splittings in semiconductor nanostructures. *Phys. Rev. B* **63**, 085310 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the State of Bavaria, the National Science Foundation and by JSPS through its FIRST programme. I.G.S. acknowledges support from the Eimskip foundation. I.A.S. acknowledges support from the 'Center of excellence in polaritronics', IRSES SPINMET and POLAPHEN projects. A.R.-I. acknowledges a German National Academic Foundation fellowship. The authors thank T. Sünner, I. Lederer and A. Schade for experimental and technical support.

Author Contributions S.H. initiated the study and guided the work together with S.R., Y.Y. and A.F. C.S., M.L. and S.H. designed and grew the wafer and performed pre-characterization. A.W. and M.K. processed the devices. A.R.-I., J.F., N.Y.K., L.W. and S.R. established an electrical/optical Fourier-space spectroscopy setup. A.R.-I., J.F., M.A., C.S., S.H., N.Y.K. and S.R. performed experiments. A.R.-I., C.S. and M.A. analysed and interpreted the experimental data, supported by S.H., S.R., V.D.K., I.G.S. and I.A.S. C.S., A.R.-I. and S.H. wrote the manuscript, with input from all co-authors. C.S. and A.R.-I. contributed equally to the study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.H. (sven.hoefling@physik.uni-wuerzburg.de) and C.S. (christian.schneider@physik.uni-wuerzburg.de).

A new anode material for oxygen evolution in molten oxide electrolysis

Antoine Allanore¹, Lan Yin^{1†} & Donald R. Sadoway¹

Molten oxide electrolysis (MOE) is an electrometallurgical technique that enables the direct production of metal in the liquid state from oxide feedstock^{1,2}, and compared with traditional methods of extractive metallurgy offers both a substantial simplification of the process and a significant reduction in energy consumption³. MOE is also considered a promising route for mitigation of CO₂ emissions in steelmaking^{3–5}, production of metals free of carbon⁶, and generation of oxygen for extra-terrestrial exploration^{7,8}. Until now, MOE has been demonstrated using anode materials that are consumable (graphite for use with ferro-alloys and titanium^{6,9}) or unaffordable for terrestrial applications (iridium for use with iron^{10,11}). To enable metal production without process carbon, MOE requires an anode material that resists depletion while sustaining oxygen evolution. The challenges for iron production are threefold. First, the process temperature is in excess of 1,538 degrees Celsius (ref. 10). Second, under anodic polarization most metals inevitably corrode in such conditions^{11–13}. Third, iron oxide undergoes spontaneous reduction on contact with most refractory metals¹⁴ and even carbon. Here we show that anodes comprising chromium-based alloys exhibit limited consumption during iron extraction and oxygen evolution by MOE. The anode stability is due to the formation of an electronically conductive solid solution of chromium(III) and aluminium oxides in the corundum structure. These findings make practicable larger-scale evaluation of MOE for the production of steel, and potentially provide a key material component enabling mitigation of greenhouse-gas emissions while producing metal of superior metallurgical quality.

Alloys of composition Cr_{1–x}Fe_x have been tested as bulk anode materials under electrochemical conditions previously demonstrated

as successful for iron extraction¹⁰. The mass fraction of iron in the alloy has been varied between 0 and 30% (that is, *x* between 0 and 0.3), the iron-rich limit corresponding to an alloy melting point equal to the process temperature (1,600 °C). The selected electrolyte composition is very basic, which is advantageous for its physico-chemical properties—for example, electrical conductivity and viscosity—but challenging in terms of anode materials compatibility¹¹.

The electrolysis cell, operated in constant-current mode (Fig. 1a) at an average anode current density exceeding 2 A cm^{–2}, exhibited a low cell voltage, which reached a steady value of around 3.8 V after slightly less than 2 h. Increasing the current density does not lead to an intolerable increase in the cell voltage, although there is an increase in its amplitude of variation. Observation of the electrolyte surface close to the anode during electrolysis confirms the evolution of gas (Supplementary Video), the generation and accumulation of which in such a small cell have a noticeable impact on the variation of voltage with time. The fact that the oxygen content of the outlet gas gradually reaches values comparable to those observed with a noble-metal anode in the same cell¹¹ is evidence of successful production of oxygen with this new anode material. Its measured faradaic efficiency—based on the recovered oxygen and taking into account typical consumption by oxidation of the inactive parts of the furnace (mainly the molybdenum current collectors)—is around 24%, similar to that previously obtained with an iridium anode. Furthermore, the oxygen concentration in the process off-gas scales with current density. The simultaneous production of metal is confirmed by the observation of the cathode after electrolysis (Fig. 1b). The iron metal, which has alloyed with the molybdenum cathode substrate, contains almost no carbon (<25 p.p.m.), a small amount of chromium (<500 p.p.m.), and an amount of

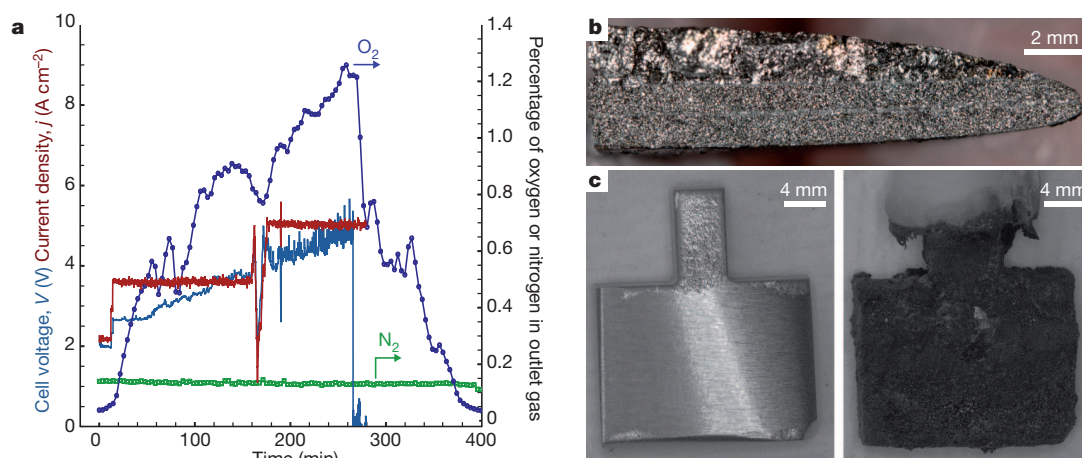


Figure 1 | Electrolysis experiments demonstrate metal and oxygen production with a macroscopically stable Cr₉₀Fe₁₀ anode. **a**, Variation of the cell voltage (left axis) and oxygen and nitrogen content of the process gas (right) during constant-current electrolysis ($T = 1,565\text{ }^{\circ}\text{C}$, $92,964\text{ C}$). **b**, Fracture of the

cathode deposit, showing the deposition of molten metal on top of the substrate. **c**, Macrographs of a Cr₉₀Fe₁₀ anode before (left) and after (right) electrolysis, showing the limited change in dimensions ($T = 1,565\text{ }^{\circ}\text{C}$, $21,923\text{ C}$).

¹Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA. [†]Present address: Department of Materials Science and Engineering, University of Illinois Urbana-Champaign, Urbana, Illinois 61801, USA.

oxygen in agreement with that predicted by thermodynamics for carbon-free iron at 1,600 °C (<4,000 p.p.m.). At the cathode the corresponding faradaic efficiency is 34%, in close agreement with previous investigations^{10,11}. Along with these promising electrochemical results, anode stability has been demonstrated (Fig. 1c): the external dimensions have barely been affected by electrolysis, as confirmed by repeated testing for various durations with a plurality of alloy compositions (Supplementary Fig. 1). The anode shows coverage by electrolyte (so-called drag-out) that has adhered to the surface.

The efficiency and stability of the $\text{Cr}_{1-x}\text{Fe}_x$ alloys for oxygen evolution is in agreement with thermodynamic calculations, which predict the solid chromium oxidation product to be stable in molten electrolyte (Supplementary Discussion A, Supplementary Fig. SA2). However, a study of the anode microstructure after electrolysis (Fig. 2a and Supplementary Fig. 2) suggests a more sophisticated mechanism. First, limited internal oxidation of the alloy is observed together with chromium depletion in the bulk metal of ~10%, effective over a distance of approximately 150 μm (Fig. 2b). Second, and of relevance to the overall anode stability, a triplex scale is observed at the alloy surface: metal oxide/mixed oxide/frozen electrolyte. Analysis by energy dispersive spectroscopy (EDS; Fig. 2a) and optical observations under polarized light (Fig. 2c) reveal that the oxidation products are found in two layers: an inner layer of chromium oxide (Cr_2O_3) containing a few per cent of aluminium, and an outer layer composed of a solid solution of Cr_2O_3 and Al_2O_3 .

The crystallographic structures of these phases have been confirmed by X-ray diffraction (XRD) analysis of the inner side of the crust (Fig. 2d, measured from the alloy side) and the powder obtained by crushing the entire crust (Fig. 2e). In the latter, a signal from the frozen

slag is also noticeable. A small amount of chromium has been found in the electrolyte close to the anode (Fig. 2a), in agreement with both the thermodynamic prediction and the experimental measurements of chromium oxide chemical solubility in an oxide melt of this specific composition (Supplementary Discussion A). The composition and morphology of the oxide scale were found to be pretty much independent of the composition of the base alloy, which influenced mainly the extent of chromium depletion (Supplementary Fig. 2). In the case of pure chromium, the inner layer was found to contain a higher concentration of aluminium, and the outer layer exhibited a non-negligible concentration of calcium.

The influence of electrolysis duration on the thickness of the chromium(III) oxide containing layer did not prove to be important (for $\text{Cr}_{90}\text{Fe}_{10}$: $75 \pm 16 \mu\text{m}$ after 2.5 h as compared to $42 \pm 38 \mu\text{m}$ after 5.2 h) on the timescale investigated, which was limited by the crucible material performance. In contrast, the initial concentration of iron in the anode was found to be of major importance: an anode containing 30% iron in the starting material proved to develop thicker chromium oxide containing layers (for $\text{Cr}_{70}\text{Fe}_{30}$: $115 \pm 26 \mu\text{m}$) and to incur higher internal depletion of chromium, leading to the formation of a low melting point alloy responsible for faster local consumption and reshaping of the anode by alloy fusion (Supplementary Figs 1 and 2).

Dry oxidation studies conducted on the same alloys in similar conditions proved that the rate of chromium oxide formation during electrolysis was comparable ($36 \pm 16 \mu\text{m}$ after 2.5h) and not sensitive to either the oxygen concentration or the alloy iron content (Supplementary Discussion B, Supplementary Figs SB2 and SB3), suggesting that electrolysis did not lead to an inescapable increase in the oxidation rate of the metal. The application of existing oxidation theory (Supplementary

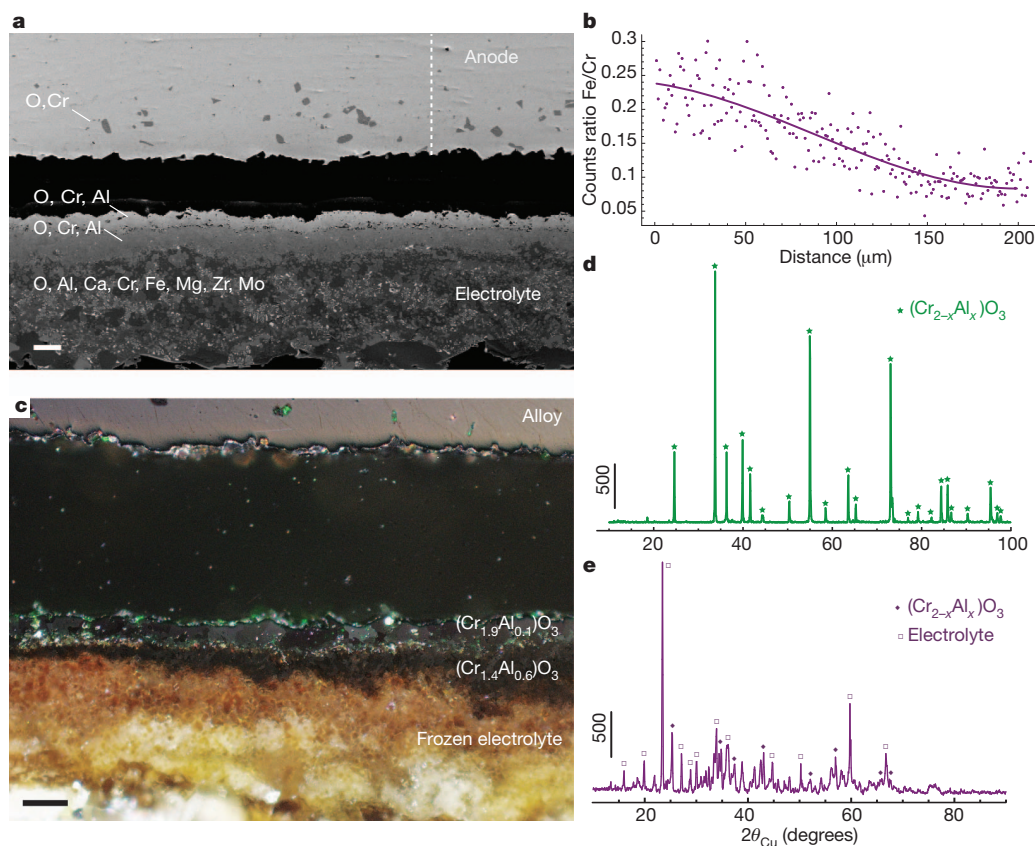


Figure 2 | Anodes developed passivation layers rich in chromium(III) oxide during electrolysis. **a**, SEM micrograph in backscattered mode of $\text{Cr}_{90}\text{Fe}_{10}$ ($T = 1,565^\circ\text{C}$, 21,920 C). Elements at left ranked by atomic concentration measured by EDS. The vertical dotted line at right corresponds to the line analysed in **b**. Scale bar, 50 μm . **b**, EDS analysis of Cr/Fe atomic ratio along the

line presented in **a**, revealing chromium depletion towards the surface. **c**, Optical micrograph of the anode/electrolyte interface with phases identified by EDS and XRD. Scale bar, 50 μm . **d**, **e**, XRD spectra of the inner side of the interface (from the alloy side; **d**) and the oxide layers in powdered form (**e**).

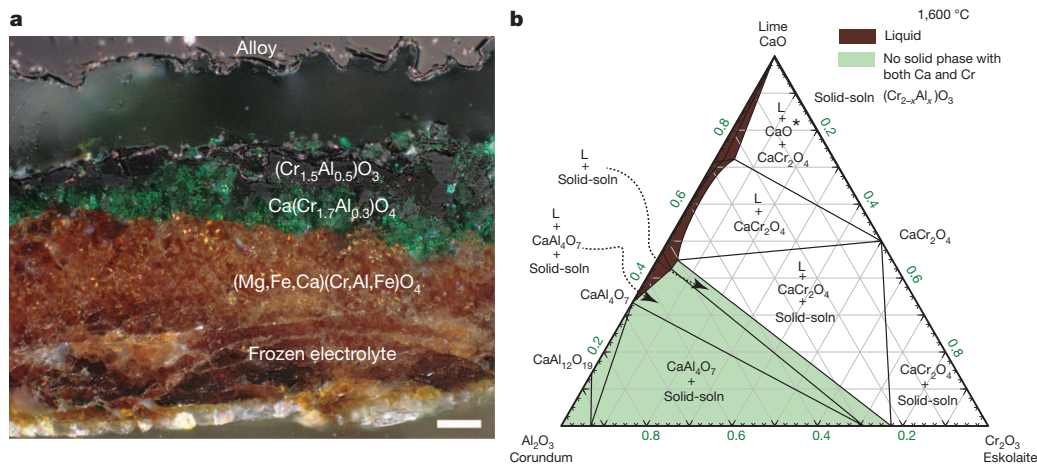


Figure 3 | $\text{Cr}_{1-x}\text{Fe}_x$ alloys immersed in electrolyte in the absence of electrolysis develop a mixed oxide layer composed of both calcium spinel and chromia-alumina solid solution. **a**, Optical micrograph of a $\text{Cr}_{90}\text{Fe}_{10}$ anode after immersion for 5 h in the electrolyte without polarization. Phase identification is based on EDS and XRD analysis. Scale bar, 50 μm . **b**, Phase

Discussion B, Supplementary Fig. SB5) suggests that in this specific electrolysis case, external polarization creates an electric field of such small magnitude across the anode scale as to have almost no influence on the oxide growth rate.

Chromium alloys immersed in the molten electrolyte in the absence of electrolysis (Fig. 3a) proved to develop a thicker layer ($98 \pm 29 \mu\text{m}$ of chromium-rich oxide layer after 5 h) than in either the electrolysis or the dry oxidation cases. Analysis of the chromium-rich oxide layers also reveals the presence of a calcium spinel phase containing chromium and aluminium (Fig. 3a and Supplementary Discussion A, Supplementary Fig. SA4).

The results obtained after electrolysis show that the oxide scale formed on the alloy is principally a solid solution of chromium(III) oxide and aluminium oxide (alumina), whereas without polarization a calcium spinel phase predominates. The phase diagram for alumina–chromium(III) oxide–calcium oxide (Fig. 3b) confirms that at low calcium oxide concentration, in green on the diagram, the stable chromium-containing solid-oxide phase is indeed the solid solution with alumina. It is likely that electro-migration under the influence of electrolysis is responsible for driving away the positive calcium ions (calcium ion is the dominant cation in the electrolyte, while aluminium is present as an aluminate anion), leading to the formation of a calcium-free solid solution of chromium and aluminium on the anode. From an oxidation perspective, the formation of the solid solution in place of the calcium spinel is seen as beneficial, because chromium diffusivity is expected to be smaller in the corundum structure than in the spinel. The diffusivity of Cr^{3+} in the chromium oxide–alumina solid solution is also predicted to be smaller than in a pure chromium oxide phase owing to the difference in lattice constant¹⁵. Finally, the formation of such a solid solution may also modify the rate-determining step for oxidation, because it has been shown that the oxygen coverage on such material is much poorer than on pure chromium(III) oxide¹⁶.

These results have two consequences. First, the successful extraction of metal and oxygen obtained with an affordable alloy-based anode material at bench-scale will enable the design of a reactor of sufficiently large scale that longer-term MOE performance can be properly assessed—for example, the specific energy consumption and cell service lifetime. This is especially important in light of the fact that the values of such performance characteristics obtained in the present laboratory-scale cell configuration are not considered to be representative of industrial conditions. The same disparity holds for other high-temperature electrolytic technologies, notably the Hall–Héroult process for extraction of aluminium¹⁷. Encouraging signs are to be found in the

diagram of chromium(III) oxide, aluminium oxide and calcium oxide at 1,600 °C, predicted from available data. Areas in brown and green represent respectively the liquid phase and the domain of immiscibility of chromium and calcium in a solid oxide phase. $\text{CrO}_2(\text{s})$ has been removed from the diagram because of its low concentration. CaO^* is in the liquid phase.

results of previous large-scale production of manganese by MOE⁶. Second, the oxidation behaviour of alloys in such an extreme environment has resulted in a novel microstructure that challenges high-temperature oxidation theory, that is, the formation of an oxide solid solution containing one element not present in the base alloy. Further work will involve the application of existing passivation theory, for example, the point defect model¹⁸, to the specific case of molten oxide electrolytes.

METHODS SUMMARY

The original set-up described elsewhere¹¹ has been adapted for the investigation of a non-noble metal anode (Methods). In summary, a vertical tube furnace constantly purged with helium is used to maintain the cell temperature. The process outlet gas composition is monitored with a gas chromatograph, which measured an average O_2 content before electrolysis of around 50 p.p.m. The cell consists in an yttria-stabilized zirconia crucible containing the oxide mixture (in wt%: 42.3 CaO , 42.3 Al_2O_3 , 5.4 MgO and 10 Fe_2O_3), pre-melted before the electrolysis experiment. All electrodes are immersed into the electrolyte from the top of the furnace, through a gas-tight cap equipped with vacuum-type feedthroughs. The horizontal cathode consists of two molybdenum disks, 3.8 cm in diameter and 0.1 cm thick, attached by nuts of the same material to a threaded molybdenum lead, 0.5 cm in diameter.

The $\text{Cr}_{1-x}\text{Fe}_x$ alloys were prepared by arc-melting high-purity metals (minimum 99.95%) of known weight. Anodes were cut to the required dimensions (typically $2 \times 1 \times 0.7 \text{ cm}$) by water-jet. The anode surfaces were ground by belt sander with no. 220 grit SiC paper before welding. Samples were connected to molybdenum leads, 0.3 cm in diameter and 90 cm long, by tungsten inert gas welding in a box flushed with argon. Evaluation of the dry oxidation features of the alloys under investigation was conducted using a thermogravimetric analyser and vertical tube furnace. Details of observation and characterization methods are provided in Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 26 October 2012; accepted 18 March 2013.

Published online 8 May 2013.

1. Aiken, R. H. Process of making iron from the ore. US patent 816, 142 (1906).
2. Esin, O. A. Electrochemistry of molten oxides. *Zh. Fiz. Khim.* **30**, 3–19 (1956).
3. Birat, J.-P., Lorrain, J.-P. & de Lassat, Y. The “ CO_2 tool”: CO_2 emissions and energy consumption of existing and breakthrough steelmaking routes. *Rev. Métal.* **106**, 325–336 (2009).
4. International Energy Agency. *Energy Technology Perspective* 179–180 (OECD/IEA, Paris, 2010); available at <http://www.iea.org/publications/freepublications/publication/etp2010.pdf>.
5. Halper, M. Stainless steel. *Time* **177**, 7–9 (2011).

6. Winand, R., Fontana, A., Segers, L., Hannaert, P. & Lacave, J. in *Int. Symp. Molten Salt Electrolysis in Metal Production* 42–50 (Institute of Mining and Metallurgy, London, 1977).
7. Oppenheim, M. J. Electrolysis of molten basalt. *Mineral. Mag.* **36**, 1104–1122 (1968).
8. Sanderson, F. How to breathe on the moon. *Nature* <http://dx.doi.org/10.1038/news.2009.803> (published online 10 August 2009).
9. Sadoway, D. R. Electrochemical processing of refractory metals. *J. Met.* **43**, 15–19 (1991).
10. Wang, D., Gmitter, A. J. & Sadoway, D. R. Production of oxygen gas and liquid metal by electrochemical decomposition of molten iron oxide. *J. Electrochem. Soc.* **158**, E51–E54 (2011).
11. Kim, H., Paramore, J., Allanore, A. & Sadoway, D. R. Electrolysis of molten iron oxide with an iridium anode: the role of electrolyte basicity. *J. Electrochem. Soc.* **158**, E101–E105 (2011).
12. Sadoway, D. R. Inert anodes for the Hall–Héroult cell: the ultimate materials challenge. *J. Met.* **53**, 34–35 (2001).
13. Di Martino, J., Rapin, C., Berthod, P., Podor, R. & Steinmetz, P. Corrosion of metals and alloys in molten glasses. Part 2: nickel and cobalt high chromium superalloys behaviour and protection. *Corros. Sci.* **46**, 1865–1881 (2004).
14. Carton, A., Rapin, C., Podor, R. & Berthod, P. Corrosion of chromium in glass melts. *J. Electrochem. Soc.* **153**, B121–B127 (2006).
15. Takahashi, N. T. *et al.* Cr diffusion in α -Al₂O₃: secondary ion mass spectroscopy and first-principles study. *Phys. Rev. B* **82**, 174302 (2010).
16. Egerton, T. A., Stone, F. S. & Vickerman, J. C. α -Cr₂O₃-Al₂O₃ solid solutions 1. The formation and stability of adsorbed oxygen. *J. Catal.* **33**, 299–306 (1974).
17. Nguyen, T. & de Nora, V. in *Light Metals* (ed. Galloway, T. J.) 385–390 (The Minerals, Metals and Materials Society, 2006).
18. MacDonald, D. The history of the point defect model for the passive state: a brief review of film growth aspects. *Electrochim. Acta* **56**, 1761–1772 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements The financial support of the American Iron and Steel Institute is acknowledged; we thank H. Kim and J. Paramore for assistance with the experimental set-up and for discussions.

Author Contributions A.A. conceived the idea and designed the study based on principles enunciated by D.R.S. A.A. and Y.L. performed the experiments, the analysis of the results, and wrote the original draft of the paper. D.R.S. edited the original manuscript and revised it for submission. All authors discussed the results and commented on the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.A. (allanore@mit.edu).

METHODS

Electrolysis experiments. The electrolyte chosen for electrolysis was composed of multiple oxides (CaO, MgO, Al_2O_3), in which 10 wt% of magnetite (Fe_3O_4) was added as iron oxide feedstock. The composition and properties of the electrolyte have been provided elsewhere¹¹. The high concentration of ionic compounds—mainly calcium oxide—is responsible for some important features of this electrolyte for electrolysis: for example, high electrical conductivity; low melting point, and low density and viscosity. The oxide powders were purchased from laboratory chemicals suppliers, with purity higher than 99 wt%. These powders were mixed in a polyethylene bottle and pre-melted in sacrificial magnesia crucibles to reduce experimental artefacts. This step has been identified as crucial for successful anode stability: if not removed before the electrolysis, impurities entrapped in the oxide compounds, such as water, were released during furnace ramping and led to contamination of the anode surface, leading to poor performance during electrolysis.

Yttria-stabilized zirconia crucibles (5 wt% Y_2O_3 , 6.6 cm diameter, 10.5 cm height, Jiangsu Province Ceramic) were used as the electrolyte container. A high-purity magnesia crucible (99.8 wt%, Ozark Technical Ceramics) was used as a secondary container to protect the tube furnace in case of failure of the inner zirconia crucible.

Chromium alloyed with 0 to 30 wt% iron ($\text{Cr}_{1-x}\text{Fe}_x$ alloy, x ranging from 0 to 0.30), has been assessed for its suitability as an oxygen-evolving anode. Strips of 1 cm thickness of these alloys were provided by Ames National Laboratory (Ames, Iowa, USA); they were obtained by arc melting in an argon atmosphere and were cut into coupons by water jet to typically $2.0 \times 1.0 \times 0.7$ cm with a top-pin of around $1.0 \times 0.5 \times 0.7$ cm for welding. The surface of the anode was polished with no. 220 SiC papers. The pin of the anode was welded to a molybdenum rod (0.3 cm diameter, 90 cm height, American Elements) that served as a current lead. Two molybdenum disks of 0.1 cm thickness and 3.8 cm diameter served as the cathode. These disks were connected to a threaded molybdenum rod by molybdenum nuts. Molybdenum leads were protected by high-purity alumina tubes, and high-purity alumina paste (Resbond 989, Cotronics Corp.) was used to protect connections in the head space above the electrolyte. The immersed part of the cathode lead was protected with an extra magnesia or zirconia tube. The anode was held vertically above the horizontal cathode disk, the corresponding projected anode area on the cathode being typically 1.4 cm^2 . This value was used in the estimation of anodic current densities.

The high-temperature pre-oxidation of the anode alloys as well as the electrolysis experiments were conducted in a vertical tube furnace fitted with lanthanum chromite heating elements (model PVT 18/100/350 with 3216P1 controller, Carbolite) and a closed-one-end alumina tube (11.4 cm inner diameter, 1 m height, 99.8%, McDanel Advanced Ceramics).

The crucible containing solid pre-melted electrolyte was placed in the hot zone of the furnace tube, the two electrodes being suspended above the crucible contents during ramping of the temperature.

High-purity helium was continuously flowed through the tube furnace at a rate of $250 \text{ cm}^3 \text{ STP min}^{-1}$ to maintain an inert atmosphere. The oxygen content of the process outlet gas was analysed by gas chromatography (CP-4900 Micro-Gas Chromatograph, Varian Inc.) and was measured to contain less than 50 p.p.m. oxygen during furnace temperature ramping and before electrolysis. The continuous monitoring of the oxygen level and the gas flow-rate at the reactor outlet allow

us to estimate the anodic faradaic efficiency specific to this laboratory configuration. It is necessary, however, to be aware of the limitation of this method, because part the oxygen is consumed by the oxidation of the metallic parts present into the tube furnace, in particular by the molybdenum current collectors. A control experiment was conducted in which a known quantity of oxygen was injected at the inlet (1% O_2 in He at $300 \text{ cm}^3 \text{ STP min}^{-1}$). The corresponding oxygen level at the reactor outlet (0.5%) suggested that around one-half of the oxygen could be consumed by internal oxidation. This 50% loss in measuring the oxygen outlet concentration has been taken into account in the anodic faradaic efficiency estimates reported in the main text. The cathode efficiency has been estimated by chemical titration (ICP-OES) of the iron content of one-quarter of the molybdenum disc cathode after electrolysis. The first set point of the furnace was $1,450^\circ\text{C}$ with a ramping rate of $1.5^\circ\text{C min}^{-1}$, followed by dwelling for 2 h. This first step allowed pre-oxidation of the anode and the formation of a protective chromium(III) oxide layer on the anode surface. The second set point was $1,600^\circ\text{C}$, the temperature at which the cathode was immersed 3 cm into the electrolyte. The anode was lowered afterwards under anodic polarization (constant cell voltage of 2 V) by 1 cm. A type-B thermocouple was also positioned above the molten electrolyte level to monitor the process temperature: $1,565^\circ\text{C}$ for a furnace set point of $1,600^\circ\text{C}$. Galvanostatic electrolysis (Argantix XDS30-500-208IF, California Instruments Corp.) was conducted at currents ranging from 2 A to 9 A for a duration ranging between 1.5 and 6 h. After electrolysis, both electrodes were raised out of the melt to a relatively cold zone in the furnace to preserve the reaction products.

Dry oxidation experiments. Oxidation experiments were carried out both in a thermogravimetric (TG) analyser (Netzsch STA 409) and the vertical tube furnace. The TG samples were $\text{Cr}_{1-x}\text{Fe}_x$ alloy beads of around 200 mg produced by arc-melting in argon atmosphere. Oxidation of $\text{Cr}_{1-x}\text{Fe}_x$ ($x = 0.1, 0.2$ and 0.3) alloys at $1,600^\circ\text{C}$ in Ar for 3 h was investigated in the tube furnace, emulating the electrolysis process. Each sample was submitted to the same preparation and experimental procedure as that for electrolysis.

Chromium(III) oxide stability experiments. The $\text{Cr}_{90}\text{Fe}_{10}$ alloy (same as that for electrolysis) and sintered Cr_2O_3 pellets were used to study the stability of Cr_2O_3 in molten electrolyte at $1,600^\circ\text{C}$ in the absence of polarization using the vertical tube furnace.

Post-experiment observations and analysis. The samples were disassembled for detailed characterization at room temperature. Part of the crust on the alloy surface was peeled off for SEM-EDS analysis (Leo VP438) on both inner and outer surfaces. Various phases on the inner surface, outer surface, and in the bulk of the crust after crushing to powder were identified by X-ray diffraction (X'Pert Pro Multipurpose Diffractometer, PANalytical). Cross-sections were obtained by mounting in thermosetting polymer, cutting with a diamond saw, and polishing with SiC papers and diamond solution down to $1 \mu\text{m}$. A high-magnification optical microscope (Olympus GX51) and SEM were used to investigate the cross-sectional surface. The carbon content of the iron product on the cathode was measured by the Leco combustion technique. Direct current plasma coupled with atomic emission spectroscopy was used to analyse the iron, chromium and molybdenum contents of the cathode deposit.

Statistics. Oxide layer thicknesses in the text are given as (mean \pm standard deviation) based on measurements at 10 locations along the layer.

Deep fracture fluids isolated in the crust since the Precambrian era

G. Holland^{1,2}, B. Sherwood Lollar³, L. Li^{3,†}, G. Lacrampe-Couloume³, G. F. Slater⁴ & C. J. Ballentine¹

Fluids trapped as inclusions within minerals can be billions of years old and preserve a record of the fluid chemistry and environment at the time of mineralization^{1–3}. Aqueous fluids that have had a similar residence time at mineral interfaces and in fractures (fracture fluids) have not been previously identified. Expulsion of fracture fluids from basement systems with low connectivity occurs through deformation and fracturing of the brittle crust⁴. The fractal nature of this process must, at some scale, preserve pockets of interconnected fluid from the earliest crustal history. In one such system, 2.8 kilometres below the surface in a South African gold mine, extant chemoautotrophic microbes have been identified in fluids isolated from the photosphere on timescales of tens of millions of years⁵. Deep fracture fluids with similar chemistry have been found in a mine in the Timmins, Ontario, area of the Canadian Precambrian Shield. Here we show that excesses of ¹²⁴Xe, ¹²⁶Xe and ¹²⁸Xe in the Timmins mine fluids can be linked to xenon isotope changes in the ancient atmosphere² and used to calculate a minimum mean residence time for this fluid of about 1.5 billion years. Further evidence of an ancient fluid system is found in ¹²⁹Xe excesses that, owing to the absence of any identifiable mantle input, are probably sourced in sediments and extracted by fluid migration processes operating during or shortly after mineralization at around 2.64 billion years ago. We also provide closed-system radiogenic noble-gas (⁴He, ²¹Ne, ⁴⁰Ar, ¹³⁶Xe) residence times. Together, the different noble gases show that ancient pockets of water can survive the crustal fracturing process and remain in the crust for billions of years.

The Precambrian crystalline basement accounts for over 30% of the exposed surface area of the continents, yet the character, communication and residence times of fluids within it remain little explored. Crystalline basement fracture waters from the Canadian, Fennoscandian and South African continental cratons have recently drawn scientific attention because they possess concentrations of H₂ as high as 7 mM (ref. 6); that is, they are as rich in H₂ as the hydrothermal vents or ocean spreading centres⁷. This H₂ has been shown to be produced by radiolysis of water⁸ and by serpentinization⁹ and to provide the energy source for deep chemoautotrophic microbial ecosystems⁵. The Witwatersrand basin in South Africa is the most extensively studied among these crustal basement systems. There, radiogenic noble gases in the deepest fracture waters provide evidence for groundwater residence times of up to 25 million years (Myr)^{5,10}, indicating that this habitat has been isolated in the crystalline basement from the photosphere on geologic timescales. How prevalent and how ancient these fluid pockets might be remains unclear.

Here we report noble-gas isotopic compositions from some of the deepest bulk fracture fluids yet sampled from a working mine (Supplementary Videos). These fluids occur 2.4 km below the surface in crystalline rocks about 2.7 billion years (Gyr) in age, from the Precambrian Shield near Timmins, Ontario, Canada. The Timmins mine is a volcanogenic massive sulphide deposit producing copper and zinc ore¹¹. Geologically, the Timmins site comprises a series of episodic volcanic

eruptions intercalated with both terrigenous and marine sedimentary deposits formed in a shallow seafloor setting about 2.719–2.711 Gyr ago¹². The deposit is relatively undeformed, metamorphosed to greenschist grade 2.67–2.69 Gyr ago¹³, with a late metasomatic event at 2.64 Gyr (ref. 14). Exploration boreholes within the mine that produce water rich in hydrogen, methane, nitrogen and helium were sampled for stable and noble-gas isotope determination, following refs 5 and 15–17 (Methods). ³He/⁴He ratios vary between 0.0052R_a and 0.0153R_a (where R_a is the ³He/⁴He ratio in air, 1.4 × 10^{−6}) confirming that radiogenic helium accounts for more than 99% of the ³He and ⁴He (ref. 18). Ne isotope data indicate no significant mantle component (Fig. 1), but rather mixing between air-saturated water (ASW) and a distinctly different radiogenic end-member, consistent with elevated

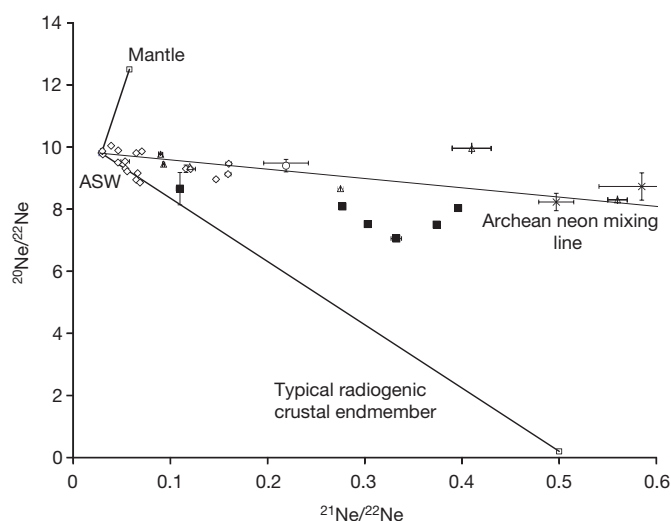


Figure 1 | Comparison of the Neon isotopic composition of free fluids and fluid inclusions in ancient crust. Data presented are: Timmins free fluids (solid squares, this work); Witwatersrand basin fracture waters (open diamonds) and fluid inclusions in bulk quartzite rock (open circles)³; fluid inclusions from vein quartz from the Kaapvaal craton, South Africa (crosses)³; CH₄-rich fluid inclusions from Yilgarn craton, Western Australia (open triangles)¹ and end-member compositions (mantle, ASW, Archean crust; and typical radiogenic crust). Nucleogenic Ne is generated by ¹⁷O, ¹⁸O(α, n)²⁰Ne, ²¹Ne, ¹⁹F(α, n)²²Na(β⁺) and ¹⁹F(α, p)²²Ne. Typical radiogenic continental environments have a ²¹Ne/²²Ne lower than would be predicted by the O/F ratio alone due to preferential siting of F proximal to U (ref. 27) and have a typical ²¹Ne/²²Ne value of around 0.47 at the ²⁰Ne/²²Ne zero intercept^{18,27}. Ne isotope data from fluid inclusions in Archean terrains^{1,3} are characterized by a significantly higher ²¹Ne/²²Ne intercept of 3.3 ± 0.2 (ref. 3), consistent with the average crustal O/F ratio¹⁸. With the exception of sample 12287-1 (see Supplementary Information), the Timmins fluids indicate ASW mixing with an Archean Ne end-member and show no evidence for radiogenic contributions or input from outside this terrain. Error bars are as shown or are smaller than the plotted symbols and are 1σ.

¹School of Earth, Atmospheric and Environmental Sciences, Manchester University, Manchester M13 9PL, UK. ²Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK. ³Department of Earth Sciences, University of Toronto, Ontario, M5S 3B1, Canada. ⁴School of Geography and Geology, McMaster University, Hamilton, Ontario L8S 4K1, Canada. [†]Present address: Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton T6G 2E3, Canada.

Table 1 | He, Ne and Ar isotope data of Timmins gas samples

Sample	$^3\text{He}/^4\text{He}$ (R_a)	Error	$^{20}\text{Ne}/^{22}\text{Ne}$	Error	$^{21}\text{Ne}/^{22}\text{Ne}$	Error	$^{40}\text{Ar}/^{36}\text{Ar}$	Error	$^{38}\text{Ar}/^{36}\text{Ar}$	Error
12261	0.0134	0.0033	7.501	0.0315	0.3739	0.0013	27173	446	0.1874	0.0005
12299-1	0.0133	0.0019	8.037	0.0657	0.3960	0.0035	44384	1554	0.1876	0.0005
12299-2	0.0153	0.0013	7.520	0.0554	0.3030	0.0032	28168	2050	0.1871	0.0004
12299-3	0.0147	0.0019	8.087	0.072	0.2770	0.0045	31787	2154	0.1875	0.0005
12287-1	0.0052	0.0012	8.660	0.5200	0.1098	0.0037	6564	123	0.1879	0.0006
12287-2	0.0113	0.0025	7.053	0.0623	0.3324	0.0054	24180	963	0.1879	0.0005

Errors quoted on all ratios are 1σ . $^3\text{He}/^4\text{He}$ are displayed relative to atmospheric ratio (R/R_a) = 1.4×10^{-6} .

$^{21}\text{Ne}/^{22}\text{Ne}$ previously identified in fluid inclusions in Archean terrains^{1,3}. One extraordinary feature of the present study is that here the Archean Ne isotope signature has been discovered not in fluids trapped in inclusions, but in the radiogenic noble-gas-rich fracture fluids discharging from exploration boreholes, with water and gas flow rates of 0.1 to over 2 litres per minute (Supplementary Table 1). $^{40}\text{Ar}/^{36}\text{Ar}$ values in these samples (up to 44,384) are the highest yet reported in a free fluid (Table 1). High-precision Xe isotope data (Table 2) using multi-collector mass spectrometry¹⁵ show isotopic excesses relative to modern atmosphere (denoted by asterisks) at all masses normalized to ^{130}Xe (Fig. 2). $^{131-136}\text{Xe}^*$ represent normal uranium fission and based on this we can rule out exotic nuclear processes or reaction routes for $^{129}\text{Xe}^*$. To our knowledge, these $^{129}\text{Xe}^*$ are the first excesses observed in a non-magmatic freely flowing fluid and show that an ancient fluid process signal has been preserved. $^{124-128}\text{Xe}^*$ appears to preserve an ancient atmosphere signal that allows a residence time to be estimated for these fluids.

The most intriguing component of the Xe mass spectra is the $^{124,126,128}\text{Xe}^*$ values (Fig. 2), indicating enrichment of the lightest isotopes of Xe. The Timmins fluids have ^{36}Ar concentrations in the original water phase similar to that of ASW (Methods) and $^{130}\text{Xe}/^{36}\text{Ar}$ ratios in excess of ASW (Supplementary Fig. 1). Xe enrichment is often inferred to be sediment-derived. However, the Xe-enriched component is not correlated with $^{124,126,128}\text{Xe}^*$, showing that the mass fractionation of $^{124-128}/^{130}\text{Xe}$ is not derived from the sediment: a finding supported by recent experimental studies¹⁹. The $^{124,126,128}\text{Xe}^*$ must therefore be a primary feature of the atmosphere-derived component.

Extreme ultraviolet radiation and asteroid bombardment in the first few hundred million years of Earth's history have often been invoked to explain the preferential loss of the light isotopes of Xe in the Earth's atmosphere relative to those of other Solar System components²⁰. Recently, evidence supporting ongoing, rather than early, Xe fractionation of the atmosphere throughout geological time has been derived from inclusions trapping ancient air in Archean barites and quartz grains^{2,21}. The evolution of the Xe isotopes in Earth's atmosphere through this mechanism can be used to identify and date ancient water that has equilibrated with the early atmosphere (Fig. 3).

If the atmosphere evolved from an initial solar composition, as indicated by Kr isotopes^{16,20}, the average Timmins $^{124,126,128}\text{Xe}^*$ value is 26%–43% of that predicted by the Pujol model², assuming the fluids

were trapped and preserved at the time of the last regional fluid event 2.64 Gyr ago. There are three explanations for this mismatch. The first is that the Pujol model has yet to be calibrated for this period and that the Timmins fluids do preserve 100% of the Xe signal of the atmosphere at this time in Earth's history—and in fact provide this calibration. The fracture fluids would then have to be 2.64 Gyr old. The second is that the Timmins fracture fluids could contain up to 74% modern Xe, and the $^{124,126,128}\text{Xe}^*$, in the context of the Pujol model, would then give a mean residence time for the fracture fluid of 1.5–1.85 Gyr (Fig. 3). Third, the accretionary Xe isotopic composition may have been chondritic and the observed $^{124,126,128}\text{Xe}^*$ would then be much closer to that expected for a 2.64-Gyr-old fluid (Fig. 3).

$^{129}\text{Xe}^*$ is also clearly resolved in the Timmins mine fluids. This could either be a magmatic signal²² or the result of the decay of ^{129}I (half-life $t_{1/2} = 15.7$ Myr). Production of ^{129}I in the crust by *in situ* ^{238}U , and more exotic radiogenic production routes of $^{129}\text{Xe}^*$, can be ruled out (Supplementary Information). We are then required to find a mantle or near-surface source of $^{129}\text{Xe}^*$. If the $^{129}\text{Xe}^*$ is mantle-derived, this should correlate with ^3He , the most unequivocal indicator of mantle input¹⁷. The absence of such a correlation and $^3\text{He}/^{129}\text{Xe}^*$ ratios in the fracture fluids ranging from 0.3 to 23, orders of magnitude lower than typical mantle values (about 500), requires extreme He/Xe elemental fractionation to preserve a Xe mantle signature without the corresponding He. We note that a mantle source for $^{129}\text{Xe}^*$ would also be associated with a $^{124}\text{Xe}^*$. However, mantle $^{129}\text{Xe}^*/^{124}\text{Xe}^*$ derived from the Bravo system¹⁷ is 1.640 ± 209 , which is an order of magnitude higher than we find in this work (199 ± 63). Therefore a mantle source for $^{129}\text{Xe}^*$ can account for only $12\% \pm 4\%$ of the $^{124-128}\text{Xe}^*$ observed.

We can next consider cosmogenic ^{129}I , which is concentrated and deposited in carbon-rich sediments²³. The periods of volcanic activity that formed the Timmins ore deposit were interspersed every few million years with carbonaceous sedimentary deposits, typically conglomerates, turbidites and greywackes^{12,14} that were subsequently isolated from the surface by the next eruptive phase. This is a scenario that may have resulted in closed-system evolution of the ^{129}I – ^{129}Xe system after each eruptive phase. Subsequent alteration of these trapped carbon-rich sediments produced the graphitic carbon observed in these sequences²⁴ and, critically, may have liberated substantial amounts of ^{129}I or ^{129}Xe . Using the average concentration of $^{129}\text{Xe}^*$ in our samples of 2.12×10^9 atoms of $^{129}\text{Xe}^*$ per g H_2O , 10 m^3 of sediment would be

Table 2 | Xe isotope data of Timmins gas samples

Sample	$^{124}\text{Xe}/^{130}\text{Xe}$	Error	$^{126}\text{Xe}/^{130}\text{Xe}$	Error	$^{128}\text{Xe}/^{130}\text{Xe}$	Error	$^{129}\text{Xe}/^{130}\text{Xe}$	Error
12261	0.023762	164	0.022071	216	0.47528	057	6.5667	171
12299-1	0.023824	203	0.022101	177	0.47454	047	6.5710	066
12299-2	0.023724	130	0.022068	132	0.47545	048	6.5730	131
12299-3	0.023895	141	0.021976	116	0.47582	048	6.5664	092
12287-1	0.023613	026	0.021924	042	0.47365	019	6.5282	026
12287-2	0.023659	135	0.022196	171	0.47417	028	6.5823	132

Sample	$^{131}\text{Xe}/^{130}\text{Xe}$	Error	$^{132}\text{Xe}/^{130}\text{Xe}$	Error	$^{134}\text{Xe}/^{130}\text{Xe}$	Error	$^{136}\text{Xe}/^{130}\text{Xe}$	Error
12261	5.2342	115	6.8712	055	3.0002	018	2.7139	114
12299-1	5.2335	063	6.8460	068	2.9630	018	2.6634	027
12299-2	5.2475	115	6.7998	190	2.8941	012	2.5774	088
12299-3	5.2424	073	6.8268	150	2.9305	023	2.6234	068
12287-1	5.2263	026	6.6445	040	2.6272	005	2.2501	018
12287-2	5.2525	094	6.8186	095	2.9206	012	2.6073	099

Errors quoted on all ratios are last significant figures at 1σ .

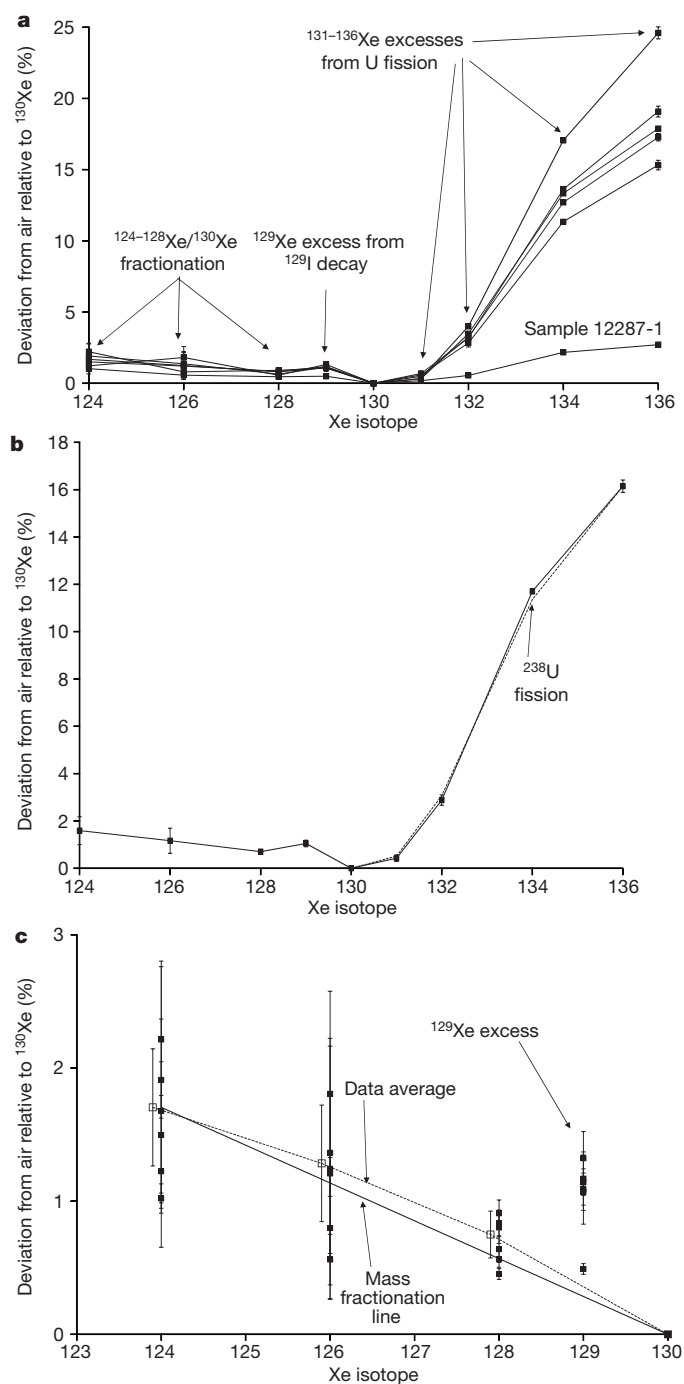


Figure 2 | Xenon isotopic spectrum. **a**, Isotopic ratios of Xe expressed as percentage deviation from air, relative to ^{130}Xe , shown for individual samples. Data show excess Xe (Xe^*) at all masses. One sample (12287-1; see Table 1) has a much larger air-derived component. This is also observed in both Ne and Ar data (Table 1), probably indicating a small amount of air contamination (Supplementary Information). **b**, $^{131,132,134,136}\text{Xe}$ is a fission product from U. The data points show the average of all samples compared to ^{238}U fission spectrum (dashed lower line). The data are clearly within error of the ^{238}U fission spectrum and suggest that no other fission source significantly contributes to the heavy xenon isotope spectrum. **c**, $^{124,126,128,129}\text{Xe}^*$ are observed in all samples. Individual samples (small solid squares) and averages of all samples except the air-contaminated sample 12287-1 are shown (open square). Errors are 1σ . Average $^{124,126,128}\text{Xe}^*$ data (dashed line) are within error of the mass fractionation line (solid line). $^{129}\text{Xe}^*$ in all samples is greater than the excess caused by mass fractionation, requiring either a cosmogenic ^{129}I or mantle ^{129}Xe source to explain most of the observed $^{129}\text{Xe}^*$ and can only be reconciled with a fluid that has preserved this component since about 2.64 Gyr (see main text).

required to provide the $^{129}\text{Xe}^*$ observed in the fracture fluids associated with 1 m^3 of Precambrian Shield rock. We note that these values are conservative and, as these processes are not well constrained, the amount of sediment required may be more than an order of magnitude less (Methods).

A sediment-derived source for the $^{129}\text{Xe}^*$ might have been initiated by regional metamorphic events that resulted in reducing hydrothermal fluids known to have penetrated the basement between 2.69 and 2.64 Gyr ago^{13,14}. Preservation of this signature in fracture fluids could only have occurred if these fluids were also from this time period. Magmatic activity, the source of any mantle-derived noble gases, also initiated around 2.79 Gyr ago and continued until around 2.64 Gyr ago^{12,14}. The region has been quiescent since this period. Therefore, regardless of whether $^{129}\text{Xe}^*$ was sediment- or mantle-derived, both hypotheses, importantly, require isolation of the sampled bulk fluids for time periods that extend over a significant period of Earth's existence.

The consequence of such an ancient isolation ought also to be seen in the radiogenic noble gases. There are two possible sources of the radiogenic noble gases in the fracture fluids: those produced *in situ* and those that are externally derived. Calculated *in situ* ages provide an important reference point (Methods). In this context the Ne isotope signature (Fig. 1) provides strong evidence that we do not have contributions to the radiogenic noble gases from anywhere other than the Archean terrain. Although it is not possible to rule out mechanisms that might concentrate radiogenic noble gases within portions of the Archean terrain, we make the a priori assumption that this is negligible to test our hypothesis of an ancient fluid. Using radionuclide U, Th and K concentrations for the local crystalline basement, we calculate the range of fluid residence times as averages of all samples (except for the air-contaminated sample 12287-1) to be $1,142 \pm 645$ Myr, $1,655 \pm 789$ Myr, $1,498 \pm 784$ Myr and $1,610 \pm 825$ Myr for ^4He ,

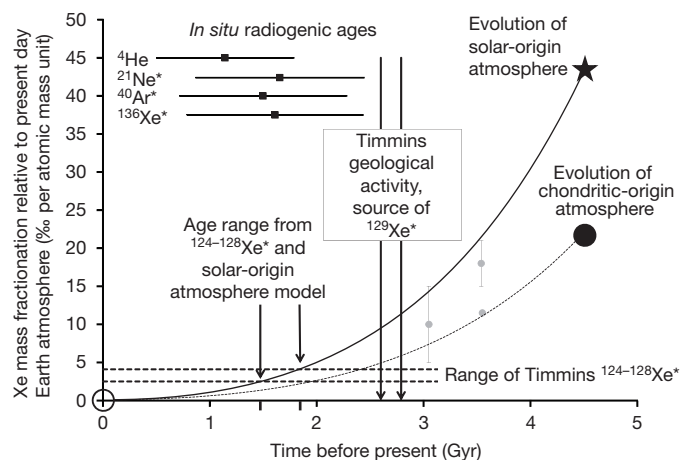


Figure 3 | The xenon isotopic evolution of Earth's atmosphere shown over time², relative to modern atmosphere. The chondritic, solar²⁸ and modern atmosphere values are shown as a solid circle, star and open circle, respectively. The solid curve is the evolution of the modern atmosphere from solar, which is our preferred initial composition¹⁶. The evolution of a chondritic-derived atmosphere is shown as a dashed curve. Grey data points are from refs 2, 21 and 29. The horizontal black dotted lines are the $\pm 1\sigma$ range of the observed $^{124,126,128}\text{Xe}^*$ of the Timmins samples (except the air-contaminated sample 12287-1), expressed per atomic mass unit (dalton). The vertical solid lines reflect the time period over which geological activity occurred at the Timmins site (approximately 2.7 to 2.64 Gyr). The solid black squares are *in situ* radiogenic noble gas ages for ^4He , ^{21}Ne , ^{40}Ar and ^{136}Xe , including errors; see Supplementary Table 3. Using a solar composition, the Timmins data require the Xe to have been isolated 1.5–1.85 Gyr ago (tick marks on x axis), or alternatively, account for 26–43% of the Xe now present if the ancient Xe component is indeed about 2.64 Gyr old. These are internally consistent with the range suggested by the ^4He , $^{21}\text{Ne}^*$, $^{40}\text{Ar}^*$ and $^{136}\text{Xe}^*$ residence times. Errors on radiogenic ages reflect one standard deviation of five samples and porosity variance (see Methods).

^{21}Ne , ^{40}Ar and ^{136}Xe respectively (see Supplementary Tables 1–3 and Methods). These are younger than 2.64 Gyr, but still exceptionally old and consistent with the range derived from the $^{124,126,128}\text{Xe}^*$ and atmosphere evolution age (Fig. 3).

$^{124-128}\text{Xe}^*$, $^{129}\text{Xe}^*$ and radiogenic noble-gas *in situ* fracture fluid residence times (Supplementary Table 3) each provide independent evidence that the sampled fracture fluids have been isolated for at least a billion years, with a ^{129}Xe signature possibly preserved for as long as 2.64 Gyr. The discovery that ancient fluid environments capable of supporting life can remain isolated for hundreds of millions to billions of years may require changes in our understanding of the extent of the Earth's crust that may be habitable, and the part that such potential buried biomes play in preserving, evolving and propagating life on planetary timescales. Mars, like the Precambrian shields on Earth, is dominated by tectonically quiescent geologic terrains which are billions of years old, some with serpentinized ultramafic rocks capable of sustaining production of reduced gases such as found in the terrestrial Archean fracture fluids^{25,26}. If such ancient fluids, with millimolar concentrations of H_2 and CH_4 , are preserved deep in the terrestrial crust on billion-year timescales, perhaps similar potential buried biomes may be preserved at depth in the subsurface of Mars on planetary timescales.

METHODS SUMMARY

Sample collection. All samples were collected at the borehole collar. A packer was placed into the opening of the borehole and sealed to the inner rock walls below the water level to seal the borehole from the mine air and minimize air contamination, following the method of refs 15 and 30. Gas and water were allowed to flow through a gas–water sampling chamber (Supplementary Videos) long enough to displace any air remaining in the borehole or the apparatus before sampling. Plastic tubing was attached to the end of the packer and the flow of gas and/or water from the borehole was directed into a graduated sampling bucket. Temperature, pH and conductivity were measured on the outflowing water from the packer by respective field kits. Water flow was measured by filling the graduated sampling vessel three times. Gas flow was measured by filling the sampling vessel with water, then displacing a known volume of water from an inverted graduated funnel. For each borehole, gas flow rates were also measured three times and average values are reported. For noble-gas analysis, gas was collected in 60-cm-long, 10-mm-diameter, internally polished refrigeration-grade copper tubing sealed using the refrigeration clamping procedure outlined by ref. 31 or crimped cold welding. This technique of gas sampling and storage is well established and ensures there has been no noble-gas contamination or loss due to leakage or diffusion.

Full Methods and any associated references are available in the online version of the paper.

Received 31 July 2012; accepted 28 March 2013.

- Kendrick, M. A., Honda, M., Walshe, J. & Petersen, K. Fluid sources and the role of abiogenic- CH_4 in Archean gold mineralization: constraints from noble gases and halogens. *Precamb. Res.* **189**, 313–327 (2011).
- Pujol, M., Marty, B. & Burgess, R. Chondritic-like xenon trapped in Archean rocks: a possible signature of the ancient atmosphere. *Earth Planet. Sci. Lett.* **308**, 298–306 (2011).
- Lippmann-Pipke, J. et al. Neon identifies two billion year old fluid component in Kaapvaal Craton. *Chem. Geol.* **283**, 287–296 (2011).
- Sleep, N. H. & Zoback, M. D. Did earthquakes keep the early crust habitable? *Astrobiology* **7**, 1023–1032 (2007).
- Lin, L.-H. et al. Long-term sustainability of a high-energy, low-diversity crustal biome. *Science* **314**, 479–482 (2006).
- Sherwood Lollar, B. et al. Isotopic signatures of CH_4 and higher hydrocarbon gases from Precambrian Shield sites: a model for abiogenic polymerization of hydrocarbons. *Geochim. Cosmochim. Acta* **72**, 4778–4795 (2008).
- Kelley, D. S. et al. A serpentine-hosted ecosystem: the Lost City hydrothermal field. *Science* **307**, 1428–1434 (2005).
- Lin, L.-H. et al. Radiolytic H_2 in continental crust: nuclear power for deep subsurface microbial communities. *Geochim. Geophys. Geosyst.* **6**, Q07003 (2005).

- Sherwood Lollar, B. et al. Abiogenic methanogenesis in crystalline rocks. *Geochim. Cosmochim. Acta* **57**, 5087–5097 (1993).
- Lippmann, J. et al. Dating ultra-deep mine waters with noble gases and ^{36}Cl . Witwatersrand Basin, South Africa. *Geochim. Cosmochim. Acta* **67**, 4597–4619 (2003).
- Bleeker, W. & Parrish, R. R. Stratigraphy and U–Pb zircon geochronology of Kidd Creek: implications for the formation of giant volcanogenic massive sulphide deposits and the tectonic history of the Abitibi greenstone belt. *Can. J. Earth Sci.* **3**, 1213–1231 (1996).
- Thurston, P. C. et al. Depositional gaps in Abitibi Greenstone Belt stratigraphy: a key to exploration for syngenetic mineralization. *Econ. Geol.* **103**, 1097–1134 (2008).
- Thompson, P. H. A new metamorphic framework for gold exploration in the Timmins-Kirkland Lake area, western Abitibi greenstone belt: Discover Abitibi Initiative. *Ontario Geologic Survey Open-File Rep.* **6162**, 1–104 (2005).
- Bleeker, W., Parrish, R. R. & Sager-Kinsman, A. In *The Giant Kidd Creek Volcanogenic Massive Sulfide Deposit, Western Abitibi Subprovince, Canada* (eds Hannington, M. D. & Barrie, C. T.) *Econ. Geol. Monogr.* **10**, 43–69 (1999).
- Sherwood Lollar, B. et al. Abiogenic formation of gaseous alkanes in the Earth's crust as a minor source of global hydrocarbon reservoirs. *Nature* **416**, 522–524 (2002).
- Holland, G., Cassidy, M. & Ballentine, C. J. Meteorite Kr in Earth's mantle suggests a late accretionary source for the atmosphere. *Science* **326**, 1522–1525 (2009).
- Holland, G. & Ballentine, C. J. Seawater subduction controls the heavy noble gas composition of the mantle. *Nature* **441**, 186–191 (2006).
- Ballentine, C. J. & Burnard, P. Production, release and transport of noble gases in the continental crust. *Rev. Mineral. Geochem.* **47**, 481–538 (2002).
- Marrocchi, Y., Robert, F. & Marty, B. Adsorption of xenon ions onto defects in organic surfaces: implications for the origin and the nature of organics in primitive meteorites. *Geochim. Cosmochim. Acta* **75**, 6255–6266 (2011).
- Pepin, R. O. On the origin and early evolution of terrestrial planet atmospheres and meteoritic volatiles. *Icarus* **92**, 2–79 (1991).
- Pujol, M., Marty, B., Burnard, P. & Philipott, P. Xenon in Archean barite: weak decay of ^{130}Ba , mass-dependent isotopic fractionation and implication for barite formation. *Geochim. Cosmochim. Acta* **73**, 6834–6846 (2009).
- Staudacher, T. Upper mantle origin for Harding County well gases. *Nature* **325**, 605–607 (1987).
- Santos, F. J. et al. ^{129}I record in a sediment core from Tinto River (Spain). *Nucl. Instrum. Methods Phys. Res. B* **259**, 503–507 (2007).
- Downes, M. J., Hodges, D. J. & Derwedumen, J. in *Gold '82: The Geology, Geochemistry and Genesis of Gold Deposits (Proc. Symp. "Gold '82". Univ. Zimbabwe, May 24–28, 1982)* (ed. Foster, R. P.) 435–448 (A. A. Balkema, 1984).
- Mumma, M. J. et al. Strong release of methane on Mars in northern summer 2003. *Science* **323**, 1041–1045 (2009).
- Ehlmann, B. L., Mustard, J. F. & Murchie, S. L. Geologic setting of serpentine-bearing rocks on Mars. *Geophys. Res. Lett.* **37**, L06201 (2010).
- Kennedy, B. M., Hiyagon, H. & Reynolds, J. Crustal neon: a striking uniformity. *Earth Planet. Sci. Lett.* **98**, 277–286 (1990).
- Wieler, R. & Baur, H. Krypton and xenon from the solar wind and solar energetic articles in two lunar ilmenites of different antiquity. *Meteoritics* **29**, 570–580 (1994).
- Srinivasan, B. Barites: anomalous xenon from spallation and neutron-induced reactions. *Earth Planet. Sci. Lett.* **31**, 129–141 (1976).
- Ward, J. A. et al. Microbial hydrocarbon gases in the Witwatersrand Basin, South Africa: implications for the deep biosphere. *Geochim. Cosmochim. Acta* **68**, 3239–3250 (2004).
- Ballentine, C. J. & Sherwood Lollar, B. Regional groundwater focusing of nitrogen and noble gases into the Hugoton–Panhandle giant gas field, USA. *Geochim. Cosmochim. Acta* **66**, 2483–2497 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Shirey for discussion on the geological history of the study areas and P. Burnard for a review. This work was funded by NSERC Discovery and CRC grants to B.S.L., a UK-NERC grant to C.J.B. and Deep Carbon Observatory support to C.J.B. and B.S.L. We are indebted to P. Calloway, P. Jurenovski, A. Marcotte and L. Kieser for assistance in sample collection.

Author Contributions G.H., B.S.L. and C.J.B. designed the project, interpreted the data and wrote the paper. G.H. analysed the samples. G.F.S. and L.L. collected the field samples. G.F.S., L.L. and G.L.-C. characterized the gas and water samples and provided critical comment and input on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.J.B. (chris.ballentine@manchester.ac.uk).

METHODS

Sample collection. All samples were collected at the borehole collar. A packer was placed into the opening of the borehole and sealed to the inner rock walls below the water level to seal the borehole from the mine air and minimize air contamination, following the method of refs 15 and 30. Gas and water were allowed to flow through a gas–water sampling chamber (Supplementary Videos) long enough to displace any air remaining in the borehole or the apparatus before sampling. Plastic tubing was attached to the end of the packer and the flow of gas and/or water from the borehole was directed into a graduated sampling bucket. Temperature, pH and conductivity were measured on the outflowing water from the packer by respective field kits. Water flow was measured by filling the graduated sampling vessel three times. Gas flow was measured by filling the sampling vessel with water, then displacing a known volume of water from an inverted graduated funnel. For each borehole, gas flow rates were also measured three times and average values are reported. For noble-gas analysis, gas was collected in 60-cm-long, 10-mm-diameter, internally polished refrigeration-grade copper tubing sealed with the refrigeration clamping procedure outlined by ref. 31 or crimped cold welding. This technique of gas sampling and storage is well established and ensures there has been no noble-gas contamination or loss due to leakage or diffusion.

^{129}I analytical method. Samples for ^{129}I analysis were filtered through 0.2- μm filters and 1-litre pre-cleaned Nalgene bottles were completely filled, leaving no head space. The bottles were then wrapped in aluminium foil and maintained in a 4 °C cold room before ^{129}I determination by accelerator mass spectrometry. Iodine in the sample was chemically reduced to iodide. 2 mg of potassium iodide carrier with known ^{129}I concentration— $^{129}\text{I}/^{127}\text{I} = (2.2 \pm 0.4) \times 10^{-14}$ —were added to the sample and the iodide was oxidized to molecular iodine by addition of sodium nitrite. The iodine was extracted into hexane and then back-extracted into a sodium bisulphite/sulphuric acid solution. Silver nitrate was added to ensure complete precipitation of the analyte as silver iodide. The dried silver iodide sample was mixed with an equal weight of 350-mesh niobium powder and pressed into stainless steel sample holders for analysis by accelerator mass spectrometry at the University of Toronto IsoTrace Facility. The sample data are normalized to IsoTrace Reference Material 2, which has been calibrated to NIST-3230 I and II standards to have $^{129}\text{I}/^{127}\text{I} = (1.313 \pm 0.017) \times 10^{-11}$. The blank (potassium iodide carrier added to de-ionized water) for this procedure is typically $(0.075 \pm 0.010) \times 10^8$ atoms per litre and the standard deviation (1σ) ranges from 5% to 10%.

Atmosphere-derived noble gases and produced water. The dominant source of atmosphere-derived noble gas isotopes in subsurface fluids is from air dissolved in groundwater (ASW) (see review³²). Oil and gas fields in contact with ancient basinal brines often show excesses of $^{84}\text{Kr}/^{36}\text{Ar}$ and $^{130}\text{Xe}/^{36}\text{Ar}$ relative to the values found in ASW. This has been attributed to sediment-derived contributions of air-derived Kr and Xe that are sorbed onto carbonaceous-rich sediments and transferred to the fluid phase through water-rock interaction. Similarly, although they are less well understood, excess $^{20}\text{Ne}/^{36}\text{Ar}$ relative to ASW are also often found in old crustal fluids^{33–35}. We use here ocean water at 10 °C as our reference ASW, containing $1.47 \times 10^{-7} \text{ cm}^3$ of ^{20}Ne , $1.01 \times 10^{-6} \text{ cm}^3$ of ^{36}Ar , $3.99 \times 10^{-8} \text{ cm}^3$ of ^{84}Kr and $5.2 \times 10^{-10} \text{ cm}^3$ of ^{130}Xe (at standard temperature and pressure, STP) per gram of H_2O (ref. 36).

All samples in this study also show excess $^{20}\text{Ne}/^{36}\text{Ar}$, $^{84}\text{Kr}/^{36}\text{Ar}$ and $^{130}\text{Xe}/^{36}\text{Ar}$ relative to ASW. A coherent increase in these values is most simply attributed to mixing between ASW and a fluid enriched in ^{20}Ne , ^{84}Kr and ^{130}Xe relative to ^{36}Ar (Supplementary Fig. 1). Importantly, there is no coherent behaviour of the radiogenic noble-gas ratios ($^4\text{He}/^{40}\text{Ar}^*$, $^{21}\text{Ne}^*/^{40}\text{Ar}^*$ and $^{136}\text{Xe}^*/^{136}\text{Xe}^*$), where the isotopic values in excess of air and denoted by the asterisk are ascribed to radiogenic sources with the atmosphere-derived noble-gas elemental ratios. From this we conclude that the ASW noble-gas patterns (Supplementary Fig. 1) must represent heterogeneity of the air-derived noble gases, incorporated into the fluids as production proceeds *in situ*, rather than an artefact of phase separation during production and sampling.

Gases are naturally exsolving from the flowing fracture waters. Gas/water volume ratios of 0.045 to 0.72 g of H_2O per cm^3 of gas (at STP) will result in near-complete transfer of any noble gases in the groundwater into the gas phase, according to Henry's Law. The concentration of the noble gases in the gas phase and the gas/water ratio is used to calculate the original concentration of the noble gases in the groundwater before exsolution. Using this method, groundwater ^{36}Ar concentrations are calculated to range between 2.23×10^{-7} and $2.46 \times 10^{-6} \text{ cm}^3$ (at STP) per gram of H_2O , with a modal value of about $8.5 \times 10^{-7} \text{ cm}^3$ (at STP) per gram of H_2O . Neglecting sample 12287-1, with its anomalously low $^{40}\text{Ar}/^{36}\text{Ar}$ ratio suggesting some degree of air contamination, the average value of this sample suite is $9.04 \times 10^{-7} \text{ cm}^3$ (at STP) per gram of H_2O and is within 9% of the reference ASW value. The similarity of the integrated *in situ* groundwater ^{36}Ar concentration with that of ASW suggests that the groundwater has undergone little, if any, open-system phase separation since last equilibrating with the atmosphere.

Radiogenic noble-gas concentrations. The variance of $^4\text{He}/^{36}\text{Ar}$ ratios in the samples is smaller than that of the water/gas ratio variability between boreholes (Supplementary Table 1). To remove the effect of sample-to-sample variance caused by differences in gas/water flow rates between boreholes, we take the average measured value (neglecting the air-contaminated sample 12287-1) of $9.04 \times 10^{-7} \text{ cm}^3$ (at STP) per gram of H_2O and use the observed $^4\text{He}/^{36}\text{Ar}$ to calculate the ^4He concentration in the same water. Neglecting sample 12287-1 again, ^4He concentrations in the *in situ* groundwater range from 0.086 to 0.204 cm^3 (at STP) per gram of H_2O (with an average of 0.136 cm^3 (at STP) per gram of H_2O). We note that any gas phase not in solution, whether under hydrostatic or lithostatic pressure, occupies negligible volume compared to the relatively incompressible water phase. At 1.0% porosity (ref. 37) and a density of 2.7 g cm^{-3} , 1 g of rock will contain 0.0037 g formation water and $5.62 \times 10^{-4} \text{ cm}^3$ (at STP) ^4He and $1.15 \times 10^{-4} \text{ cm}^3$ (at STP) $^{40}\text{Ar}^*$. A similar calculation is made for $^{21}\text{Ne}^*$ and $^{136}\text{Xe}^*$.

Porosity range. The porosity of the basement rocks hosting the sampled fluids nevertheless represents an uncertainty in deriving the bulk-rock radiogenic noble-gas content. Bucher and Stober³⁷ report a characteristic porosity for basement rocks of 1.0%. Well tests give flow effective porosities in the Black Forest basement of 0.1% to 2.1% (ref. 38), with a corresponding range of low hydraulic conductivities for tests on rocks from the Canadian Shield³⁹. Reported porosity ranges in granite are similar and range between 0.3% and 1.0% (ref. 40) and 0.9% and 2.3% (ref. 41). In gneiss the data are more limited, but the Urach test site in Germany reports a fracture porosity of 0.1%–0.5% (ref. 42). We adopt the value of 1.0% from ref. 37 and assign a 1σ uncertainty to this of $\pm 0.45\%$ to include the observed porosity range at a 2σ level of confidence.

External versus internal sources of radiogenic gases. There are two possible sources of the radiogenic noble gases in the fracture fluids: those produced *in situ* and those that are externally derived. Within basinal sedimentary aquifers, for example, a significant additional source is derived from the deeper crust and accumulates beneath the aquitard⁴³. The mechanism of transfer of noble gases from the crystalline basement to shallow regions is uncertain, with evidence for episodic, advective⁴⁴ and fault-controlled⁴⁵ mechanisms as well as uniform continual input^{34,46}. While the contribution of external sources of radiogenic noble gases cannot be ruled out, the limits of the '*in situ*' noble-gas ages must be recognized and additional evidence for their validity acquired. This might be the geological context or data, such as presented here, on xenon light-isotope deviations from atmosphere and ^{129}Xe excess values that support ancient sedimentary sources and processes.

***In situ* fluid residence time.** *In situ* ^4He and ^{40}Ar production rates can be calculated using:

$$^4\text{He} = 8 \times [^{238}\text{U}] \times (e^{\lambda_{238}t} - 1) + 7 \times [^{235}\text{U}] \times (e^{\lambda_{235}t} - 1) + 6 \times [^{232}\text{Th}] \times (e^{\lambda_{232}t} - 1) \quad (1)$$

and

$$^{40}\text{Ar} = 0.105 \times [^{40}\text{K}] \times (e^{\lambda_{40}t} - 1) \quad (2)$$

where ^4He and ^{40}Ar are the radiogenic noble-gas concentrations that would be produced within a rock in time t , containing $[^{238}\text{U}]$, $[^{235}\text{U}]$, $[^{232}\text{Th}]$ and $[^{40}\text{K}]$ concentrations respectively. λ_{238} , λ_{235} , λ_{232} and λ_{40} are the decay constants for the respective radio-isotopes and the value 0.105 is the fraction of ^{40}K that decays to ^{40}Ar . $^{21}\text{Ne}^*$ and $^{136}\text{Xe}^*$ are produced in proportion to radiogenic helium with $^4\text{He}/^{21}\text{Ne}^*$ and $^4\text{He}/^{136}\text{Xe}^*$ production ratios of 9.96×10^6 and 3.033×10^8 respectively¹⁸. Observed $^4\text{He}/^{21}\text{Ne}^*$ in this study's samples are between 0.54 and 0.97 (average, 0.725) of the crustal production ratio, whereas observed $^4\text{He}/^{136}\text{Xe}^*$ are 0.54–0.936 (average, 0.743) of the crustal production ratio. The $^{21}\text{Ne}^*/^{136}\text{Xe}^*$ ratios are between 0.769 and 1.14 (average, 0.975) of the crustal production ratio. We note that it seems that the ^4He concentration is approximately 30% lower than that expected from the $^{21}\text{Ne}^*$ and $^{136}\text{Xe}^*$. The $^{21}\text{Ne}^*/^{136}\text{Xe}^*$ values are within the expected production range and the small variance from sample to sample in all radiogenic ratios provides further evidence that there are no significant artefacts caused by degassing kinetics or phase fractionation during gas production or sampling.

The resolved radiogenic $^4\text{He}/^{40}\text{Ar}^*$ ratio in all samples ranges from 3.99 to 5.36 (average, 4.68). This is a very small range and falls between the predicted ratios of 6.56 and 3.55 for production at 0 Gyr and 2.67 Gyr ago respectively within the Kidd-Munro group, probably representing the average composition of the parent elements (Th = 9 parts per million (p.p.m.), U = 2 p.p.m. and K = 2%; ref. 47). We also consider the regional values for the southwest Abitibi belt (Th = 4.31 p.p.m., U = 0.91 p.p.m. and K = 1.48%; ref. 48), from which we calculate $^4\text{He}/^{40}\text{Ar}^*$ ratios of 4.14 and 2.24 for 0 and 2.67 Gyr ago, respectively. We note that there is better agreement between predicted and measured $^4\text{He}/^{40}\text{Ar}^*$ using Kidd-Munro-group

concentrations of parent elements and therefore provide *in situ* ages in the manuscript using these values. Because Kidd-Munro-group parent-nuclide concentrations are higher than other compositions (for example, southwest Abitibi Belt), *in situ* residence times presented here are a conservative estimate yielding lower limits for the ages. Using equations (1) and (2) and the radio-element concentrations for the Kidd-Munro group, we calculate ^4He accumulation ages of 1.14×10^9 years and ^{40}Ar accumulation ages of 1.50×10^9 years, respectively (Supplementary Table 3). This method assumes all release coefficients $\lambda_{\text{He,Ne,Ar,Xe}} = 1$, which is consistent with the observed radiogenic $^4\text{He}/^{21}\text{Ne}^*/^{40}\text{Ar}^*/^{136}\text{Xe}^*$ at or close to the predicted production ratios and again yields lower *in situ* age limits.

Calculating $^{129}\text{Xe}^*$ concentrations. Measured radiogenic $^{129}\text{Xe}^*$ concentrations in the six individual samples range from $1.09\text{--}3.2 \times 10^{-11} \text{ cm}^3$ of $^{129}\text{Xe}^*$ per cm^3 of gas. As with radiogenic noble gases, we assume the co-produced water is completely degassed but originally contained ^{36}Ar with near-ASW concentrations of $9.04 \times 10^{-7} \text{ cm}^3$ (at STP) per gram of H_2O . ^{130}Xe concentration in fracture water is therefore derived from observed $^{130}\text{Xe}/^{36}\text{Ar}$ and ASW ^{36}Ar concentration. Once ^{130}Xe concentration is obtained, $^{129}\text{Xe}^*$ is calculated from the increase in $^{129}\text{Xe}/^{130}\text{Xe}$ over the value in air. Multiplying these values by $(6.02 \times 10^{23}/22,400)$ converts these initial concentrations in the fracture fluid to atoms per gram of H_2O . Undegassed samples therefore contain $1.24\text{--}3.61 \times 10^9$ atoms $^{129}\text{Xe}^*$ per gram of H_2O , with an average of 2.85×10^9 atoms of $^{129}\text{Xe}^*$ per gram of H_2O . Correcting for the mass fractionation observed in $^{124}\text{--}^{128}\text{Xe}/^{130}\text{Xe}$ for each sample reduces this to $(0.90\text{--}2.70) \times 10^9$ atoms of $^{129}\text{Xe}^*$ per gram of H_2O , with an average of 2.12×10^9 atoms of $^{129}\text{Xe}^*$ per gram of H_2O .

***In situ* production routes for $^{129}\text{Xe}^*$.** *In situ* ^{238}U production of ^{129}I in the crust typically reaches a steady state after about 80 Myr in crystalline basement hosted waters⁴⁹. The ^{129}I concentration in sample 12299-2, determined by AMS, is 1.23×10^8 atoms per litre. At steady state this would require 5.48×10^{11} years to produce the observed 2.12×10^{12} atoms of $^{129}\text{Xe}^*$ per litre of H_2O . Given that this is greater than the age of the Universe, this allows us to discount *in situ* ^{238}U -derived ^{129}I as a significant source of the $^{129}\text{Xe}^*$. Radiogenic sources of ^{129}Xe , such as epithermal or thermal neutron capture by tellurium, or chemically fractionated fission products from natural reactors, all produce significant and distinct quantities of heavy Xe isotopes¹⁸. However, the $^{131}\text{--}^{136}\text{Xe}/^{130}\text{Xe}$ pattern is indistinguishable from simple ^{238}U fission (Fig. 2b), and these exotic production routes can also be discounted.

Calculation of cosmogenic ^{129}I contribution from sediment. Cosmic-ray spallation of xenon in the atmosphere produces an equilibrium cosmogenic ^{129}I concentration today of about 2×10^9 atoms per litre in sediment porewater⁵⁰. This is three orders of magnitude too low to account for the observed $^{129}\text{Xe}^*$. Nevertheless, iodine is sequestered by organic sediments, with pre-nuclear fallout sediment cores documented with ^{129}I concentrations of 1×10^6 atoms per gram of sediment (ref. 23). We assess here whether the iodine concentration in the sediments associated with the Timmins mineralizing system is sufficient to generate the $^{129}\text{Xe}^*$ observed in the Timmins mine fluid samples. Assuming $1\% \pm 0.45\%$ porosity, 1 m^3 of rock would contain $10,000 \pm 4,500 \text{ g}$ of H_2O in fluid fractures. Assuming total degassing of water, the average $^{129}\text{Xe}^*$ from six measured boreholes is 2.12×10^9 atoms of $^{129}\text{Xe}^*$ per g of H_2O . Based on this, 1 m^3 of Precambrian Shield contains $2.12 \times 10^9 \times (10,000 \pm 4,500 \text{ g}) = (2.12 \pm 0.95) \times 10^{13}$ atoms of $^{129}\text{Xe}^*$. Overlying sediment containing 1×10^6 atoms per gram of ^{129}I and a density of 2 g cm^{-3} will produce 2×10^{12} atoms of $^{129}\text{Xe}^*$ per m^3 . Approximately 10 m^3 of sediment is therefore required to supply the $^{129}\text{Xe}^*$ observed in the fracture fluids associated with 1 m^3 of Precambrian Shield rock. This mechanism requires 100% efficient mobilization of iodine (if mobilization occurs before ^{129}I decay) and xenon (if mobilization occurs after ^{129}I decay), and retention of all iodine (or xenon) during compaction and dewatering.

However, we note that there is significant uncertainty in these calculations: (1) Ref. 51 reports an ^{129}I concentration of 1.4×10^7 atoms per gram of marine sediments: an order of magnitude higher than used in the calculation above. (2) The elevated $^{124}\text{Xe}/^{130}\text{Xe}$ observed in these samples, if it is an indication of a less fractionated ancient atmosphere, would require 30% more xenon to have been present in the atmosphere at the time of formation²⁰, generating enhanced production of ^{129}I . (3) The ultraviolet flux penetrating the anoxic Archean atmosphere was considerably higher⁵². (4) The intensity of ultraviolet radiation was likely to have been greater from the younger Sun⁵³. These factors may combine to reduce the amount of sediment required by more than two orders of magnitude from 10 m^3 to less than 0.1 m^3 for each 1 m^3 of 1% porosity rock.

Geological setting. The stratiform massive base metal sulphide (Cu–Zn) deposits of the Timmins mine lie on the western end of the Abitibi greenstone province of the Canadian Shield^{12,14}. The area is characterized by steeply dipping Archean

felsic, mafic and ultramafic deposits interlayered with metasedimentary deposits referred to as the Kidd-Munro assemblage. Deposition began with massive felsic rhyolite and rhyolitic volcanoclastic rocks extruded onto ancient sea bed deposits (ultramafics) in a proximal seafloor setting. The volcanoclastic deposits appear to be from brecciation and resedimentation of a previously consolidated rhyolite unit. Within these units are flow-banded subvolcanic sills and dike intrusions of massive rhyolite and metavolcanoclastic rhyolites⁵⁴. Stringer ore is closely associated with the upper section of the felsic piles, probably forming as a result of silica-rich and metal-rich hydrothermal fluids circulating just below the sea floor. Banded and massive sulphide ores overlie the stringer zones, deposited on the volcanics as precipitate where hydrothermal solutions rich in metals entered the sea water⁵⁴. Discontinuous carbonaceous horizons ranging from argillite to chert occur within the volcanoclastics, are closely associated with the zones of ore deposition, and are believed to represent periods of quiet chemical sedimentation intervals of 2–27 Myr between periods of episodic volcanism¹². The deposit is relatively undeformed, metamorphosed to greenschist grade between 2.67 to 2.69 Gyr, with a late metasomatic event at 2.64 Gyr (refs 12 and 14).

32. Schwarzenbach, R. P., Gschwend, P. M. & Imboden, D. M. *Environmental Organic Chemistry* (John Wiley and Sons, 1993).
33. Torgersen, T. & Kennedy, B. M. Air-Xe enrichments in Elk Hills oil field gases: role of water in migration and storage. *Earth Planet. Sci. Lett.* **167**, 239–253 (1999).
34. Torgersen, T., Kennedy, B. M. & van Soest, M. C. Diffusive separation of noble gases and noble gas abundance patterns in sedimentary rocks. *Earth Planet. Sci. Lett.* **226**, 477–489 (2004).
35. Zhou, Z., Ballentine, C. J., Schoell, M. & Stevens, S. H. Identifying and quantifying natural CO_2 sequestration processes over geological timescales: the Jackson Dome CO_2 deposit, USA. *Geochim. Cosmochim. Acta* **86**, 257–275 (2012).
36. Kipfer, R., Aeschbach-Hertig, W., Peeters, F. & Stute, M. in *Noble Gases in Geochemistry and Cosmochemistry* (eds Porcelli, D., Ballentine, C. J. & Wieler, R.) *Rev. Mineral. Geochem.* **47**, 615–700 (2002).
37. Bucher, K. & Stober, I. Fluids in the upper continental crust. *Geofluids* **10**, 241–253 (2010).
38. Stober, I. Permeabilities and chemical properties of water in crystalline rocks of the Black Forest, Germany. *Aquat. Geochem.* **3**, 43–60 (1997).
39. Stober, I. & Bucher, K. Hydraulic properties of the crystalline basement. *Hydrogeol. J.* **15**, 213–224 (2007).
40. Guillot, L. *et al.* Porosity changes in a granite close to quarry faces: quantification and distribution by ^{14}C -MMA and Hg porosimetry. *Eur. Phys. J. A* **9**, 137–146 (2000).
41. Aquilina, L., de Dreuz, J. R., Bour, O. & Davy, P. Porosity and fluid velocities in the upper continental crust (2 to 4 km) inferred from injection tests at the Soultz-sous-Forêts geothermal site. *Geochim. Cosmochim. Acta* **68**, 2405–2415 (2004).
42. Stober, I. & Bucher, K. in *Hydrogeology of Crystalline Rocks* (eds Stober, I. & Bucher, K.) 53–78 (Kluwer, 2000).
43. Torgersen, T. & Clarke, W. B. Helium accumulation in groundwater I: an evaluation of sources and the continental flux of crustal ^4He in the Great Artesian Basin, Australia. *Geochim. Cosmochim. Acta* **49**, 1211–1218 (1985).
44. Ballentine, C. J., O'Nions, R. K., Oxburgh, E. R., Horvath, F. & Deak, J. Rare gas constraints on hydrocarbon accumulation, crustal degassing and groundwater flow in the Pannonian Basin. *Earth Planet. Sci. Lett.* **105**, 229–246 (1991).
45. Kulongoski, J. T., Hilton, D. R. & Izibicki, J. A. Helium isotope studies in the Mojave Desert, California: Implications for groundwater chronology and regional seismicity. *Chem. Geol.* **202**, 95–113 (2003).
46. Bethke, C. M., Zhao, X. & Torgersen, T. Groundwater flow and the He-4 distribution in the Great Artesian Basin of Australia. *J. Geophys. Res. Solid Earth* **104**, 12999–13011 (1999).
47. Ketchum, J. W. F. *et al.* Pericontinental Crustal Growth of the Southwestern Abitibi Subprovince, Canada—U–Pb, Hf, and Nd Isotope Evidence. *Econ. Geol.* **103**, 1151–1184 (2008).
48. Moulton, B. J. A. *et al.* Archean subaqueous high-silica rhyolite coulees: examples from the Kidd-Munro Assemblage in the Abitibi Subprovince. *Precamb. Res.* **189**, 389–403 (2011).
49. Bottomley, D. J., Renaud, R., Kotzer, T. & Clark, I. D. Iodine-129 constraints on residence times of deep marine brines in the Canadian Shield. *Geology* **30**, 587–590 (2002).
50. Martin, J. B., Gieskes, J. M., Torres, M. & Kastner, M. Bromine and iodine in Peru margin sediments and pore fluids: implications for fluid origins. *Geochim. Cosmochim. Acta* **57**, 4377–4389 (1993).
51. Price, N. B. & Calvert, S. E. The geochemistry of iodine in oxidised and reduced Recent marine sediments. *Geochim. Cosmochim. Acta* **37**, 2149–2158 (1973).
52. Farquhar, J., Bao, H. M. & Thieme, M. Atmospheric influence of Earth's earliest sulfur cycle. *Science* **289**, 756–758 (2000).
53. Ribas, I. *et al.* Evolution of the solar activity over time and effects on planetary atmospheres. II. κ^1 Ceti, an analog of the Sun when life arose on Earth. *Astrophys. J.* **714**, 384–395 (2010).
54. Walker, R. R., Matulich, A., Amos, A. C., Watkins, J. J. & Mannard, G. W. The geology of the Kidd Creek mine. *Econ. Geol.* **70**, 80–89 (1975).

Adaptive dynamics under development-based genotype–phenotype maps

Isaac Salazar-Ciudad^{1,2} & Miquel Marín-Riera²

It is not known whether natural selection can encounter any given phenotype that can be produced by genetic variation. There has been a long-lasting debate about the processes that limit adaptation^{1–6} and, consequently, about how well adapted phenotypes are. Here we examine how development may affect adaptation, by decomposing the genotype–fitness map—the association between each genotype and its fitness—into two: one mapping genotype to phenotype by means of a computational model of organ development⁷, and one mapping phenotype to fitness. In the map of phenotype and fitness, the fitness of each individual is based on the similarity between realized morphology and optimal morphology. We use three different simulations to map phenotype to fitness, and these differ in the way in which similarity is calculated: similarity is calculated for each trait (in terms of each cell position individually), for a large or a small number of phenotypic landmarks (the ‘many-traits’ and ‘few-traits’ phenotype–fitness maps), and by measuring the overall surface roughness of morphology (the ‘roughness’ phenotype–fitness map). Evolution is simulated by applying the genotype–phenotype map and one phenotype–fitness map to each individual in the population, as well as random mutation and drift. We show that the complexity of the genotype–phenotype map prevents substantial adaptation in some of the phenotype–fitness maps: sustained adaptation is only possible using ‘roughness’ or ‘few-traits’ phenotype–fitness maps. The results contribute developmental understanding to the long-standing question of which aspects of phenotype can be effectively optimized by natural selection.

The relationship between genetic and phenotypic variation, or the genotype–phenotype map^{2,8}, has been argued to be complex at all levels of organization at which it has been studied^{4,9–11}. If this relationship is complex, small genetic differences between a parent and a descendant do not necessarily lead to small phenotypic differences between them. In other words, individuals with genotypes that produce adapted morphologies may not have any offspring with better-adapted morphologies. If so, this would mean that evolution by natural selection would be likely to often become trapped in local optima^{8,12} (morphologies that are not the most adaptive, and from which it is not possible to mutate to more adaptive morphologies).

There are a number of models of genotype–phenotype maps and adaptive dynamics^{13–15}. For morphological phenotypes, understanding the genotype–phenotype map should also provide an understanding of the mechanisms of development^{4,16}. Development has been argued repeatedly to constrain adaptation, by affecting which morphological variations are possible and which are not, and depending on the complexity of the genotype–phenotype map that it entails^{4,11}. It seems clear that development has an effect on adaptive dynamics but it is not evident how strong that effect is. Here we explore how a realistic approximation of this genotype–phenotype map provides a richer and more quantitative understanding of the limitations that development may impose on adaptation.

We use a computational model of tooth development as an example of a genotype–phenotype map for the morphology of a complex organ.

The tooth model summarizes mathematically the basic genetic and cellular interactions that regulate tooth-shape development⁷. The strength of those interactions is encoded by the values of the model parameters, and we take those values as a proxy for individual genotypes (see Supplementary Information). The developmental model produces a phenotype, a three-dimensional morphology, from the parameters of each individual. As a result of model dynamics, teeth with different parameter values can have different numbers of cells and different numbers of identifiable morphological landmarks, such as tooth cusps. The model has been used previously to reproduce morphological variation at the level of natural populations⁷, and thus represents current understanding of how development leads to adult morphology and variation in this morphology (a genotype–phenotype map). We use this model to derive information about the complexity of the genotype–phenotype maps of a representative complex functional anatomical feature.

We assume, for simplicity, that individual fitness is determined exclusively by the morphology arising from the developmental model. We use three kinds of phenotype–fitness maps, or natural-selection criteria, to explore how—and how much—adaptation occurs, given our developmental genotype–phenotype map. These criteria assign fitnesses to individuals based on three different methods of measuring the distance between each individual’s morphology and a pre-determined optimal morphology. Thus, the evolutionary process is simulated by applying these two mappings, genotype–morphology and morphology–fitness, to each individual. Each individual’s fitness determines its chances of contributing to the next generation. The simulations also include random mutation and drift. Mutation is implemented as changes in the developmental model parameters (see Supplementary Information).

The first selection criterion (Fig. 1a), the Euclidean morphological distance (EMD), considers all cell-level traits in each tooth: this is the position of each cell in each three-dimensional tooth form. This is all the morphological resolution provided by the model: all other measurable traits are derived from these cell-level traits. For the EMD criterion, the value of each cell-level trait in a form is compared against the value of the corresponding trait in an optimal morphology. These two morphologies can have different numbers of traits (see Methods). The fitness of each individual is then the sum of the fitness contribution of each cell-level trait.

The landmark-based selection criteria (the ‘few-traits’ and the ‘many-traits’ criteria) (Fig. 1b) are similar to the EMD criterion but based only on a set number of traits, the spatial coordinates of landmarks shared by all morphologies that are considered. Landmarks are chosen among the cusps present in teeth (see Methods). Different criteria use different numbers of traits: the ‘few-traits’ criterion uses 2 or 4 traits, and the ‘many-traits’ criterion uses 8 or 13.

For the orientation patch count (OPC) selection criterion (Fig. 1c), the high-level¹⁷ overall ruggedness of the tooth surface is measured as the number of facets in the morphologies with different orientations in the *x–y* plane (Fig. 1) (see Methods). Thus, fitness is not assigned by

¹Evolutionary phenomics group. Developmental Biology Program, Institute of Biotechnology, University of Helsinki, PO Box 56, FIN-00014 Helsinki, Finland. ²Genomics, Bioinformatics and Evolution Group. Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona, Cerdanyola del Vallès 08193, Spain.

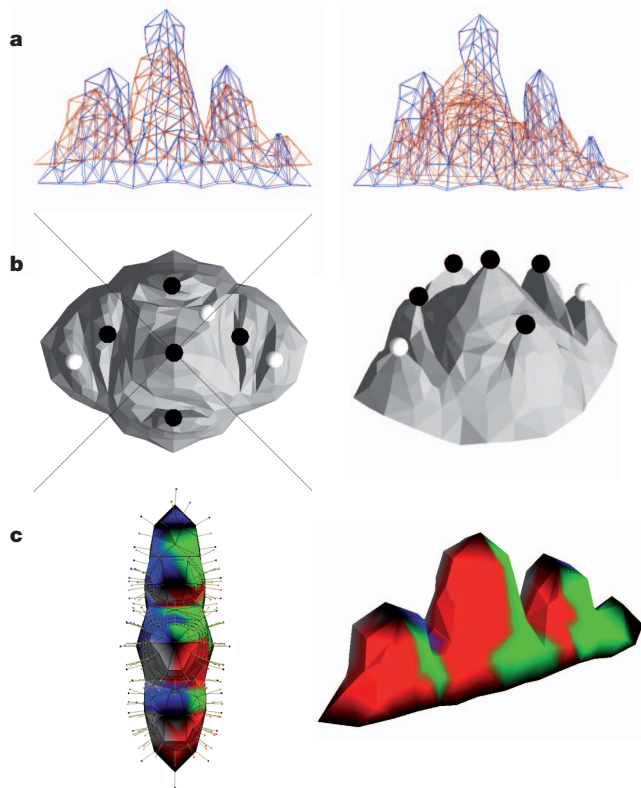


Figure 1 | Natural-selection criteria used. **a**, Morphologies assigned by the EMD criterion; the sum of the distances of each cell in one morphology to the closest cell in the other morphology, divided by the surfaces of the two morphologies. Two pairs of morphologies are shown; smaller EMD morphologies are shown to the left and larger EMD morphologies are shown to the right, with optimal morphologies shown in red and individuals' morphologies shown in blue, in both cases. Left, EMD = 56%, right, EMD = 92%. **b**, The landmarks criterion; the coordinates of five cusps (black dots) on each morphology are taken as traits. We take the central cusp and the tallest cusp in each quadrant (quadrant separated by a cross). White dots are shorter cusps that are not measured. The distances between these and the corresponding traits in the optimum morphology are measured. **c**, The OPC criterion; sets of adjacent points with equal slope orientation are grouped (groups with the same slope orientation are shown in the same colour). The number of these groups is the OPC. The tooth shown in the figure has a OPC of 12. The lines show the normal vector to each cell surface; dots at the end of the lines show the colour of that cell's group. Left and right images show the same teeth in different views.

comparing single traits between a morphology and the optimum morphology but by comparing each trait with its neighbouring cell-level traits and comparing the resulting OPC with that of the optimum.

These three phenotype–fitness maps can be seen as lying on an idealized spectrum of possibilities from all traits being selected, to some traits being selected and to no traits being selected. They also represent three different views, in the literature, on phenotypic evolution. Optimality theory proposes that natural selection is the dominant force in evolution and that, consequently, most¹⁷ phenotypes should be expected to be optimal (or at least to be optimal more often than not¹⁸). The EMD criterion is an extreme version of the idealized view that most phenotypes are optimal and that this can be studied by decomposing phenotypes into single traits that are themselves optimal^{3,17}. A milder version of this view is to select for a large number of landmark-based traits to be optimal: the many-traits selection criterion. In fact, morphologic adaptation has been traditionally understood by looking at a limited number of characters^{19,20} and landmarks²¹. The OPC¹⁶ and other related measures²² are alternative ways to understand adaptations that are not directly related to specific landmarks while still considering all cell-level traits. More generally, it has been suggested

that phenotype–fitness maps should be degenerate, or many-to-one, in the sense that many different phenotypes should have the same fitness²³. Thus, we chose to implement OPC because it is a degenerate measure that has been used to study teeth previously^{16,24,25}.

Each evolutionary simulation started with all individuals having the same randomly chosen initial tooth morphology. The optimal morphology in each simulation was chosen to be at a morphological distance of 20%, 40%, 60% or 80% from the initial morphology (for each of the selection criterion; see Supplementary Information).

The simulations show that, initially, absolute fitness increases rapidly by relatively large steps but over generations these steps become smaller and less frequent (see Supplementary Fig. 2). Individuals with fitness assigned by the EMD criterion do not reach optimum tooth morphology often and most do not exhibit substantial adaptation (fitness does not increase much over evolutionary time (Fig. 2 and Supplementary Fig. 3). This is the case even in simulations in which the initial and optimal teeth are only one mutation away. When the population is large (10,000), the EMD criterion does in some cases reach the optimum morphology but only when that optimum is at a phenotypic distance of 40% or less from the initial morphology in the evolutionary simulation. This population size is on the same order of magnitude as the estimated population sizes for mammals²⁶. Furthermore, morphologies at a 20% or 40% EMD from each other (Supplementary Figs 4 and 5) are barely distinguishable but cannot transition to each other (unless very large populations are considered). Individuals with fitness assigned by the many-traits criterion seldom reach the optimum (less than 40% of simulations for populations of 1,000), but show substantial adaptation on average, particularly in large populations. However, progressively higher levels of adaptation are attained for fewer trait-selection criteria (see also Supplementary Fig. 3).

Our results (Fig. 2c) indicate clearly that adaptation for the EMD decreases rapidly with the phenotypic distance between initial and

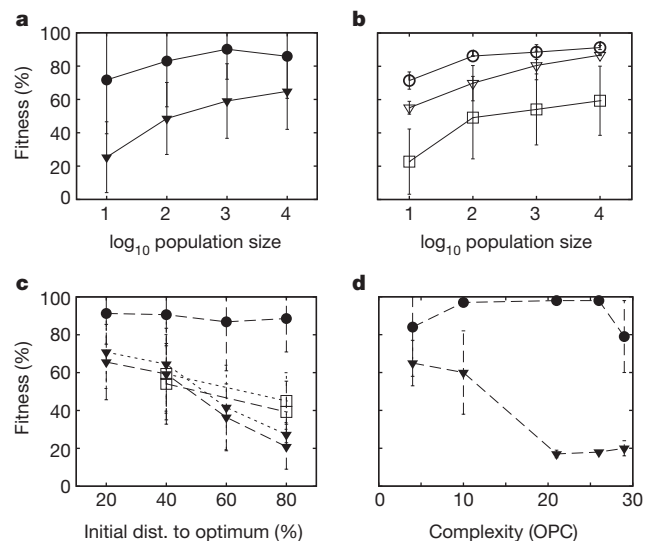


Figure 2 | Fitness assigned by the EMD criterion is modest, and the optimum morphology is rarely reached. **a**, Average \pm s.d. of the fitness reached in EMD ($n = 345$) and OPC ($n = 309$ for increasing OPC, $n = 309$ for decreasing OPC) simulations versus population size (starting at 40% distance from the optimum). **b**, The same as **a** for principal components analysis (PCA, $n = 64$) and non-PCA ($n = 193$) landmark criteria. **c**, Average final fitness in EMD and OPC simulations versus starting distances to the optimum and population sizes ($n = 154$ for EMD, $P = 1,000$; $n = 60$ for EMD, $P = 10,000$; $n = 47$ for OPC, $P = 1,000$). **d**, Average final fitness for EMD and OPC simulations, with initial morphologies of different OPC ($n = 63$). Filled triangles correspond to EMD simulations, filled circles correspond to OPCs, open triangles and circles to many traits and few traits, respectively, and open squares to PCA based simulations. In **c** and **d**, dashed lines indicate simulations run for a population size of 1,000 and dotted lines indicate simulations run for a population size of 10,000.

optimal morphologies (Fig. 2c). The EMD only leads to substantial amounts of adaptation when the initial and optimal teeth are very similar or when both are simple unicuspid teeth (Fig. 2d). In this case there is only a small amount of adaptive morphological change. In contrast, high degrees of adaptation are reached with the OPC criterion, even when the distances between initial and optimal teeth are large (Fig. 2). In fact, fitness assigned using the OPC selection criterion often reaches the optimum morphology, and leads to cumulative morphological changes that are larger than the ones reached when the EMD and many-traits selection criteria are used (Supplementary Fig. 7c).

Our interpretation of the results is that, for complex genotype–phenotype maps, adaptation can only occur if the overall genotype–fitness map is sufficiently simple. Adaptation is possible under the OPC criterion because it is degenerate; that is, the same OPC values are found in forms that differ in their morphological details¹⁶. In Fig. 3 we have quantified the degeneracy in each phenotype–fitness map (see Supplementary Figs 8, 9, 10 and 11 for related measures). In degenerate maps, such as the OPC map, genetically related individuals can have quite different phenotypes, because of the complex genotype–phenotype map, but can still have the same fitness. In addition, a large range of OPC values can be found among the offspring of most individuals so, as the figures indicate, most adaptive phenotypes can be reached through small adaptive changes. In this way, populations do not get easily trapped in local adaptive peaks (that is, in local optima).

The EMD and landmark phenotype–fitness maps are simple and smooth: there is a single peak phenotype and all other phenotypes have a fitness smoothly proportional to the distance to that phenotype (Fig. 3, and Supplementary Figs 8, 9, 10 and 11). However, as these figures indicate, degeneracy is low and, consequently, there is no simplifying effect over the overall genotype–fitness map. Degeneracy is low because it is defined at the level of the traits that contribute to fitness. This pertains to all traits for OPC and EMD, many traits for the many-traits criteria and few traits for the few-traits landmark-based criteria. The few-traits criterion works better for adaptive evolution simply

because fewer traits need to be optimized. However, this occurs only with a complex genotype–phenotype map: with a simple genotype–phenotype map the optimum morphology has been shown to be reached irrespective of the number of traits²⁷. With a complex genotype–phenotype map, optimizing many traits implies moving further in the phenotypic space and increasing the chances of becoming trapped in a local peak.

In the case of teeth, the OPC has been shown to correlate with diet in rodents, carnivorans, primates and bats^{16,24,25}; low and high OPC values correspond to animal and plant diets, respectively. However, our understanding of evolutionary adaptive transitions in morphology has traditionally been based on detailed analysis of specific characters^{19,20} and landmarks²¹. Our suggestion is that the correlation between the OPC and diet may not necessarily arise from specific OPC values being the optimal morphological solution to specific diets. It is possible that the OPC happens to be the adaptive solution, among several possible

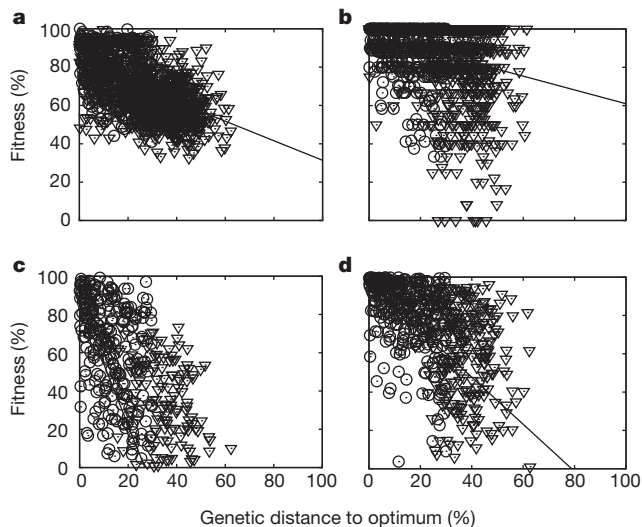


Figure 3 | Degeneracy of phenotype–fitness maps. Many mutant offspring were obtained from the same initial individuals as in Fig. 2. **a–d**, Fitness of each mutant phenotype versus genetic distance between parent and mutant was assigned using the EMD criterion (**a**) the OPC criterion (**b**), the many-traits landmark criterion (**c**, 13 traits) or few-traits landmark criterion (**d**, 2 traits). Linear regression equations were $y = -0.5871x + 87.70$ (**a**, $n = 699$); $y = -0.3763x + 95.17$ (**b**, $n = 699$); $y = 3.206x - 0.1429$ (**c**, $n = 242$); $y = 1.199x + 95.68$ (**d**, $n = 242$). The EMD, many-traits and few-traits landmark-based criteria have relatively well correlated landscapes ($r^2 = 0.3549$, and $r^2 = 0.2632$ and 0.1897 , respectively) but the OPC criterion does not ($r^2 = 0.06082$). Circles correspond to mutants with one change in the parameters and triangles to mutants with five changes.

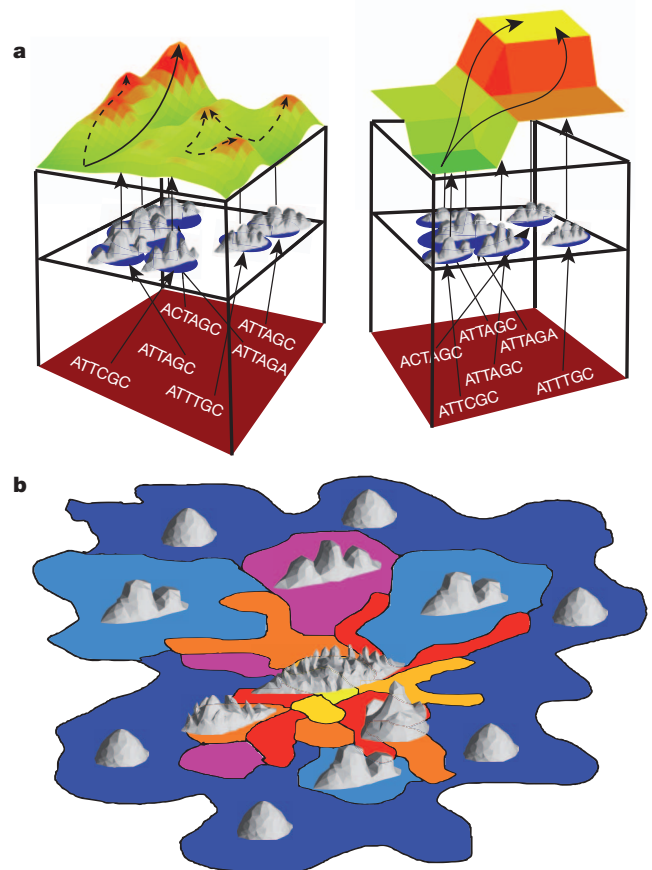


Figure 4 | Conceptual interpretation of the decomposition of the genotype–fitness map. **a**, The layers represent (from bottom to top), genetic variation space, morphological variation space and fitness space. An adaptive landscape is the direct mapping between the genetic space and fitness. Arrows represent mappings between genotype and morphology (development) and between morphology and fitness (selection criteria) (Fig. 3b, and Supplementary Figs 8, 9, 10, 11 and 12). Note that the lines between the genetic space and the morphological space cross each other. This represents how similar genotypes can lead to rather different morphologies. **b**, Idealized two-dimensional representation of the tooth model parameter space. Different colours represent hypervolumes of the parameter space with resulting morphologies of similar complexity. Blue, low complexity (dark blue are hypervolumes of higher complexity than the light blue); pink/magenta and red, intermediate complexity (pink/magenta are hypervolumes of higher complexity than the red); orange and yellow, high complexity (orange are hypervolumes of higher complexity than the yellow). The hypervolumes of complex morphologies are small but are in contact with many regions of different complexity (see Supplementary Fig. 12b).

solutions, as it has a degenerate phenotype–fitness map that allows substantial adaptation despite the complexity of the genotype–phenotype map.

In general, our results suggest that natural selection cannot find adaptive morphologies in which most, or even many, of their traits are adaptive because real genotype–phenotype maps are too complex. Our analysis does not take ecology specifically into consideration: it is possible that in certain environments, maximal fitness requires a very specific morphology characterized by a unique combination of many trait values. However, our results suggest that this is only achievable if the initial phenotypes in the population are very similar to the optimal one (Fig. 2, and Supplementary Figs 4 and 5). In contrast, substantial sustained morphological adaptation may have required degenerate phenotype–fitness maps, as in the OPC, or may have occurred only with respect to a small subset of the traits, as in the few-traits criterion. Our results should apply even when the selection criteria change over time or when no selection occurs for long periods of time. In these cases, adaptation should still occur mostly during the time periods in which the OPC or the few-traits phenotype–fitness maps are in place.

Our results do not preclude natural selection from having a crucial effect on most morphological traits. Specific selective responses of most traits in a morphology could occur because of correlations with few traits being directly adaptive^{4,5} or because of selection on the overall roughness, or similarly degenerate properties, of morphology. This degeneracy may also facilitate the evolution (in the specific case of teeth) of upper and lower occlusion without affecting the OPC²⁸. We propose that the diversity of forms in evolution is partly the result of the degenerate phenotype–fitness maps providing multiple solutions to the same problems.

Degenerate maps have been found in models of RNA folding and evolution⁹. In these, the degeneracy occurs in the genotype–phenotype map and not in the phenotype–fitness map (some level of degeneracy occurs also in our genotype–phenotype map).

There has been some controversy in the study of adaptive landscapes, concerning how rugged or uncorrelated these landscapes are. Studies based on gene networks tend to view such landscapes as uncorrelated¹², whereas some studies of population genetics assume that those landscapes need to be correlated for adaptation to be possible²⁹. Our results present a potential point of connection between these two views. The genotype–phenotype map would indeed be complex or rugged (as our current understanding of development indicates), but the phenotype–fitness maps would be simple or degenerate, as in the few-traits and OPC criteria (Fig. 4). This effectively leads to an adaptive landscape that is rugged, but not too rugged for adaptation to occur.

METHODS SUMMARY

This article uses a computational model of evolution that is based on a developmental computational model. This model produces three-dimensional morphologies based on genotypes represented by the values of the development-model parameters, and on a set of selection criteria that act on those morphologies to determine which individuals pass to the next generation. The evolutionary model is implemented in a population of individuals over generations, with selection and mutation based on the parameters of the developmental model. The developmental model is described in detail in the publication in which it was first used⁹. In the Supplementary Information, we describe in detail the how the evolutionary model uses the developmental model. The evolutionary model is simply used to apply mutation, based on the model parameters, to individuals in the population, at a rate μ (this determines the developmental model for each individual); to apply one of the three selection criteria on the resulting phenotype (morphology) to determine its fitness; and to determine which individuals contribute to the next generation based on that fitness (the methods section describes these criteria in detail). Mutation is implemented by adding or subtracting a proportion of the value of a randomly chosen parameter, in a random individual. Selection is implemented by choosing each individual in a generation stochastically from individuals in the previous generation and making the probability of being chosen proportional to individual relative fitness.

Full Methods and any associated references are available in the online version of the paper.

Received 13 September 2012; accepted 2 April 2013.

Published online 1 May 2013.

- Haldane, J. B. S. *The Causes of Evolution* (Harper Brothers, 1932).
- Wright, S. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. VI Intern. Congress Genet.* **1**, 356–366 (1932).
- Gould, S. J. & Lewontin, R. C. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist program. *Proc. R. Soc. Lond. B* **205**, 581–598 (1979).
- Alberch, P. In *Evolution and Development* Dahlem Konferenzen (ed. Bonner, J. T.) 313–332 (Springer, 1982).
- Lande, R. & Arnold, S. J. The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226 (1983).
- Fontana, W. Modelling 'evo-devo' with RNA. *Bioessays* **24**, 1164–1177 (2002).
- Salazar-Ciudad, I. & Jernvall, J. A computational model of teeth and the developmental origins of morphological variation. *Nature* **464**, 583–586 (2010).
- Waddington, C. H. *The Strategy of the Genes* (George Allen and Unwin, 1957).
- Huynen, M. A., Stadler, P. F. & Fontana, W. Smoothness within ruggedness: the role of neutrality in adaptation. *Proc. Natl Acad. Sci. USA* **93**, 397–401 (1996).
- Ferrada, E. & Wagner, A. A comparison of genotype–phenotype maps for RNA and proteins. *Biophys. J.* **102**, 1916–1925 (2012).
- Salazar-Ciudad, I. Developmental constraints vs. variational properties: how pattern formation can help to understand evolution and development. *J. Exp. Zool.* **306B**, 107–125 (2006).
- Kauffman, S. A. *The Origins of Order* (Oxford Univ. Press, 1993).
- Wagner, A. Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization. *Proc. Natl Acad. Sci. USA* **91**, 4387–4391 (1994).
- Hansen, T. F. & Wagner, G. P. Modeling genetic architecture: a multilinear theory of gene interaction. *Theor. Popul. Biol.* **59**, 61–86 (2001).
- Salazar-Ciudad, I. & Jernvall, J. How different types of pattern formation mechanisms affect the evolution of form and development. *Evol. Dev.* **6**, 6–16 (2004).
- Evans, A. R. *et al.* High-level similarity of dentitions in carnivorans and rodents. *Nature* **445**, 78–81 (2007).
- Orzack, S. H. & Sober, E. Optimality models and the test of adaptationism. *Am. Nat.* **143**, 361–380 (1994).
- Maynard Smith, J. Optimization theory in evolution. *Ann. Rev. Ecol. Syst.* **9**, 31–56 (1978).
- Ji, Q. *et al.* The earliest known eutherian mammal. *Nature* **416**, 816–822 (2002).
- Charles, C. *et al.* Regulation of tooth number by fine-tuning levels of receptor-tyrosine kinase signaling. *Development* **138**, 4063–4073 (2011).
- Klingenberg, C. P. Morphometrics and the role of the phenotype in studies of the evolution of developmental mechanisms. *Gene* **287**, 3–10 (2002).
- Bunn, J. M. *et al.* Comparing Dirichlet normal surface energy of tooth crowns, a new technique of molar shape quantification for dietary inference, with previous methods in isolation and in combination. *Am. J. Phys. Anthropol.* **145**, 247–261 (2011).
- Alfaro, M. E. *et al.* Evolutionary consequences of many-to-one mapping of jaw morphology to mechanics in labrid fishes. *Am. Nat.* **165**, E140–E154 (2005).
- Santana, S. E., Strait, S. & Dumont, E. R. The better to eat you with: functional correlates of tooth structure in bats. *Funct. Ecol.* **25**, 839–847 (2011).
- Godfrey, L. R. *et al.* Dental topography indicates ecological contraction of lemur communities. *Am. J. Phys. Anthropol.* **148**, 215–227 (2012).
- Charlesworth, B. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nature Rev. Genet.* **10**, 195–205 (2009).
- Tenaillon, O. *et al.* Quantifying organismal complexity using a population genetic approach. *PLoS ONE* **2**, e217 (2007).
- Polly, P. D., Le Comber, S. C. & Burland, T. M. On the occlusal fit of tribosphenic molars: Are we underestimating species diversity in the Mesozoic? *J. Mamm. Evol.* **12**, 283–299 (2005).
- Orr, H. A. The population genetics of adaptation on correlated fitness landscapes: the block model. *Evolution* **60**, 1113–1124 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Jernvall, M. Brun Usan, I. Salvador Martinez, A. Matamoros, S. Newman and R. Zimm for comments and the CSC (IT Center for Science). This research was funded by the Finnish Academy (WBS 1250271) and the Spanish Ministry of Science and Innovation (BFU2010-17044).

Author Contributions I.S.-C. conceived the study; and M.M.-R. and I.S.-C. constructed the evolutionary model. M.M.-R. carried out computer simulations and quantitative analyses. I.S.-C. and M.M.-R. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to I.S.-C. (isaac.salazar@uab.cat).

METHODS

EMD phenotype–fitness map. EMD is a measure of the disparity between two forms or morphologies (see Fig. 1a). This measure considers all cell-level traits (each cell position in three dimensions). It is defined as the mean distance of one cell in one morphology to the closest cell in the other morphology, corrected by size:

$$D = \frac{\sum_{k=1}^{n_1} d_{\min 12k} + \sum_{k=1}^{n_2} d_{\min 21k}}{n_1 + n_2} \quad (1)$$

where D is the EMD between two morphologies (such as an individual's morphology and the optimal morphology in a simulation), $d_{\min 12k}$ is the Euclidean distance, in three-dimensional space, between cell k at morphology 1 and the cell in morphology 2 that is closest to that cell (again in three-dimensional space; min denotes minimum (the distance to the closest cell)) and vice versa for $d_{\min 21k}$. n_1 and n_2 are the number of cells in morphology 1 and 2, respectively. Notice that this phenotypic distance measurement can be applied to morphologies made of different number of cells. This is important because in the model, different genotypes lead to morphologies with different number of cells and different numbers of morphological features (such as cusps). In each comparison, morphologies are re-scaled and rotated by a Procrustes algorithm, so that the final configuration has the minimal EMD possible between the two morphologies. To correct for size and to make the distance dimensionless, we divide the distance by the square root of the sum of the surface areas of the two morphologies:

$$d_{\text{EMD}} = \frac{D}{\sqrt{(S_1 + S_2)}} \quad (2)$$

where d_{EMD} is the final phenotypic distance, D is the distance resulting from the Procrustes algorithm, and S_1 and S_2 are the surface areas of morphology 1 and 2, respectively.

The d_{EMD} resulting from comparing an individual's morphology to an optimal morphology is used to calculate the absolute fitness of that individual.

Landmark-distance phenotype–fitness map. Landmarks are taken at tooth cusps. Therefore, this criterion can only be applied to teeth with a sufficient number of cusps (see Supplementary Fig. 13, type 4). The first landmark is the height of the highest cusp. The teeth into four quadrants (anterior, posterior, buccal and lingual) (see Fig. 1b). The second, third, fourth and fifth landmarks are identified as the tallest cusps in the anterior, posterior, buccal and lingual quadrants, respectively. Teeth with fewer cusps than required for measuring a given number of traits were given 0 absolute fitness. The traits could be chosen by two criteria; raw landmarks and PCA-based traits.

For raw landmarks, we identified up to 13 quantitative traits as the components (x , y and z) of the position vectors of the five landmarks. Tooth developmental dynamics in the model ensure that the highest cusp forms in the centre (it is also the first cusp to form). By taking this position as the point (0,0,0), we ensure that all model morphologies are in the same spatial reference system. For the central cusp, only the height is considered, because its x and y positions never change as a result of the tooth model dynamics.

Simulations were run for a different number of traits (n_t) (2, 4, 8 and 13 traits). The phenotypic distance measured using landmark criteria, d_{lan} , from a tooth to its optimum is calculated as the square root of the sum of differences between each trait and the trait's optimal value:

$$d_{\text{lan}} = \sqrt{\sum_{k=1}^{n_t} (c_k - o_k)^2} \quad (3)$$

where n_t is the number of traits under selection, c_k is the trait value of the k trait and o_k is the optimal value of the k trait. This measure is simply the distance between two points in a n_t -dimensional morphospace.

We also defined the traits as the scores of the principal components resulting from a PCA of variations of the landmarks. We calculate the PCA in a subset of the morphological space by taking the initial phenotypes and generating 300 random mutants (changing one to three parameter values per individual) and using them as a population sample to perform the PCA. Note that this PCA is only representative of a small hypervolume of the parameter space (around the initial morphologies). It is not representative of the variation possible from the model in general (that is too vast, complex and multidimensional to be efficiently characterized by a PCA). Note also that this method cannot be applied to the whole morphology, as the PCA on these can only use data sets of the same dimensionality (that is, the same number of landmarks), and different morphologies in the sample population often have different numbers of cells and different numbers of cusps. Thus, we took five different landmarks (13 traits) and carried out the PCA to get the principal components of variation. From the 13 possible principal components, we only took into account those that represented a notable amount of variation (>1%). There were 8 principal components in this group. In the evolutionary simulations, the principal-component loadings were used to calculate the scores of a given phenotype on the coordinate system of the principal components. The phenotypic distance for a given phenotype was calculated in the same way as the raw landmark-based distances, but using the principal-component scores as trait values and comparing them with the optimal PC scores (as in equation (3)). This way, selection acts always, and independently, on all of the principal components. The PCA was carried out using PAST software (<http://folk.uio.no/ohammer/past/>).

OPC phenotype–fitness map. OPC is a multivariate measure of tooth-surface complexity¹⁷ based on the number of differently oriented slopes on a morphology. Each cell-level trait on a surface is classified by the orientation of the projection of its normal vector (the vector normal to the tooth surface in that cell point) on the x - y plane, then all adjacent cells with the same orientation are grouped into patches. The count of all patches on the morphology surface is the OPC value. We established four orientation classes corresponding to the four quadrants of the x - y plane (see Fig. 1c). A smoothing algorithm was applied onto the normal vectors to erase random noise from the morphology:

$$\vec{n}'_i = \vec{n}_i + \lambda \sum_{j=1}^V \vec{n}_j \quad (4)$$

where \vec{n}'_i is the smoothed normal vector, \vec{n}_i is the original normal vector, λ is a parameter of the smoothing algorithm, V is the total number of neighbour cells and \vec{n}_j is the normal vector of a neighbour cell. We always ran the smoothing algorithm five times per morphology with $\lambda = 0.1$.

The phenotypic distance by the OPC criterion, d_{OPC} , is calculated as the absolute value of the relative difference between the OPC value of a morphology and the optimal OPC in each simulation:

$$d_{\text{OPC}} = \left(\frac{c - c_{\text{opt}}}{c_0} \right) \quad (5)$$

where c is the OPC value of a morphology, c_{opt} is the optimal OPC value and c_0 is the OPC value of the morphology at the beginning of the simulation.

Signature of ocean warming in global fisheries catch

William W. L. Cheung¹, Reg Watson² & Daniel Pauly³

Marine fishes and invertebrates respond to ocean warming through distribution shifts, generally to higher latitudes and deeper waters. Consequently, fisheries should be affected by 'tropicalization' of catch^{1–4} (increasing dominance of warm-water species). However, a signature of such climate-change effects on global fisheries catch has so far not been detected. Here we report such an index, the mean temperature of the catch (MTC), that is calculated from the average inferred temperature preference of exploited species weighted by their annual catch. Our results show that, after accounting for the effects of fishing and large-scale oceanographic variability, global MTC increased at a rate of 0.19 degrees Celsius per decade between 1970 and 2006, and non-tropical MTC increased at a rate of 0.23 degrees Celsius per decade. In tropical areas, MTC increased initially because of the reduction in the proportion of subtropical species catches, but subsequently stabilized as scope for further tropicalization of communities became limited. Changes in MTC in 52 large marine ecosystems, covering the majority of the world's coastal and shelf areas, are significantly and positively related to regional changes in sea surface temperature⁵. This study shows that ocean warming has already affected global fisheries in the past four decades, highlighting the immediate need to develop adaptation plans to minimize the effect of such warming on the economy and food security of coastal communities, particularly in tropical regions^{6,7}.

Assessing the effects of climate change on marine fisheries is one of the challenges for sustainable management of marine ecosystems. Marine biota respond to ocean warming through changes in distributions and abundance^{3,4,8}, phenology⁹ and body size¹⁰, leading to alteration of community structure^{4,8} and trophic interactions¹¹, and ultimately affecting fisheries⁷. Studies assessing the potential effects of climate change on global fisheries under scenarios of greenhouse gas emissions predict a large-scale redistribution of maximum fisheries catch potential^{12,13}

and increased vulnerability of many coastal fisheries to climate change, particularly in the tropics⁶. Climate change effects on some fisheries have been detected^{14,15}. For example, the rapid increase in catches of red mullet (*Mullus barbatus*), a warm-water species, around the UK is suggested to be related to ocean warming¹⁵. However, a signature of the effect of climate change on global fisheries has so far not been demonstrated. Because marine fisheries contribute to the economy and food security of many coastal communities, fisheries' responses to climate change need to be better understood to inform the development of effective management and adaptation policies⁷.

Shifts in distributions of exploited stocks are expected to affect their availability to fisheries. Spatial distributions of marine fishes and invertebrates are strongly dependent on the relationship between physiological optima and limits under different temperatures, oxygen levels and other biotic and abiotic conditions^{16,17}. Organisms living in temperatures outside their thermal optima experience reduced aerobic scope, negatively affecting their growth and reproduction, and ultimately reducing their abundance¹⁸. In contrast, some species may find that areas previously unsuitable for their survival have become more favourable. Consequently, the distribution margins and centroids of many marine fishes and invertebrates shift following changes in ocean conditions¹⁶, with a tropicalization of species compositions occurring as the ocean warms¹. Concomitantly, species with limited dispersal potential or a narrow range of temperature tolerance, such as in semi-enclosed seas and tropics, will decrease in overall abundance^{18,19}.

We represent the temperature preference of species in fisheries catch by an index called mean temperature of the catch (MTC) (Fig. 1). This global index builds on previous use of temperature preferences to evaluate potential effects of climate change on fisheries locally^{14,20}. We propose that ocean warming leads to increased catches of warmer-water species and decreased catches of colder-water species, resulting in an

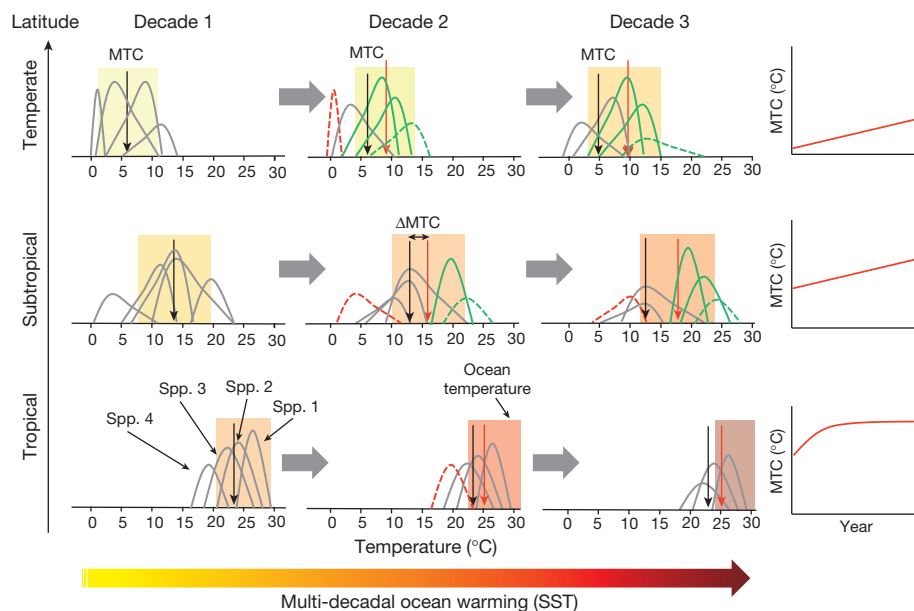


Figure 1 | Changes in catch species composition in relation to ocean warming and the resulting changes in MTC. Species distributions are related to ocean temperature (coloured bars) and temperature preferences of the exploited species (grey curves). Increase and decrease in abundance due to ocean warming are indicated by green curves and the reduction in area under the grey curves, respectively. The vertical black and red arrows represent MTC in the initial and subsequent decades, respectively. Δ MTC represents the difference in MTC relative to the initial decade. Species local extinction and invasion because of warming are indicated by red and green dotted curves, respectively. The expected changes in MTC over time are shown on the right.

¹Changing Ocean Research Unit, Fisheries Centre, The University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. ²Institute for Marine and Antarctic Studies, University of Tasmania, Taroona, Tasmania 7001, Australia. ³Sea Around Us Project, The University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada.

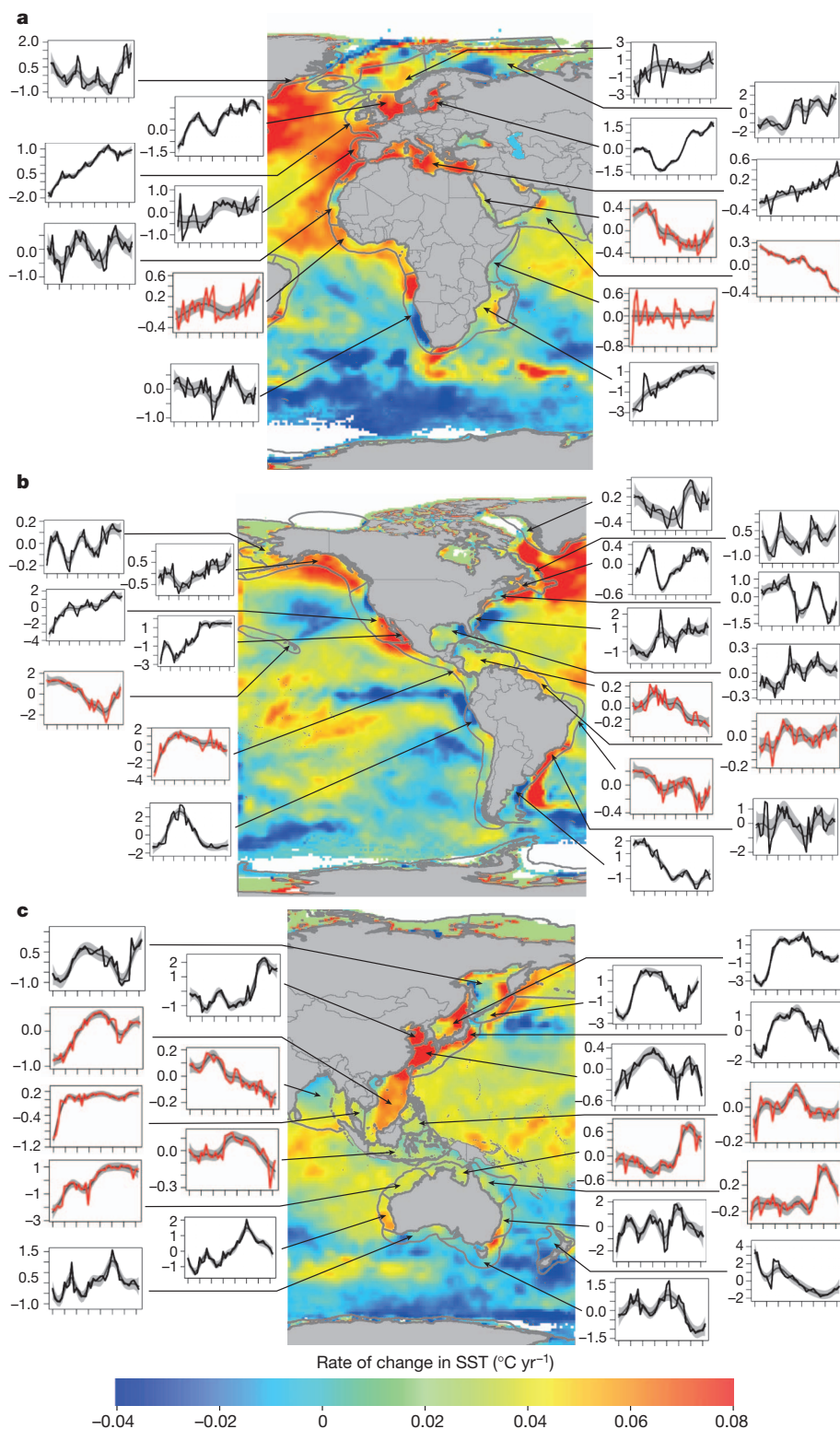


Figure 2 | Changes in MTC and SST of 52 LMEs between 1970 and 2006. Change over time (*x* axis; years between 1970 and 2006) in MTC anomalies relative to the mean of the time series (*y* axis; °C). Rate of SST change is shown by the colour scale. **a**, Eurasia; **b**, the Americas; **c**, Asia and Oceania. Red lines represent tropical LMEs. To highlight the nonlinear trends of MTC anomalies in this figure, each MTC time series was fitted with a spline smoothing function using GAMM: grey line, mean; shaded area, 95% confidence interval.

increase in MTC. We further propose that the general relationship between MTC and sea surface temperature (SST) is modified in the tropics, where catch compositions consist of both tropical and subtropical species. Warming would cause the initial decrease in the proportion of catches of subtropical species, resulting in increases in MTC because of the poleward shift in the equatorial range boundary of subtropical species. Such initial increases in MTC would then stabilize as catches become dominated by tropical species, but the scope for further tropicalization is limited¹⁶. Further warming, to levels that

exceed the temperature preference and tolerance of tropical species, is then expected to reduce their abundance¹² without changes in MTC.

We calculated MTC as the average of the inferred temperature preference of exploited species (990 species in total) weighted by their annual catch between 1970 and 2006, for 52 large marine ecosystems (LMEs) that account for most of the world's fisheries (Methods). To attribute changes in MTC to ocean warming, we used a generalized additive mixed model (GAMM) with an ensemble of different combinations of factors to account for the potential effects of large-scale

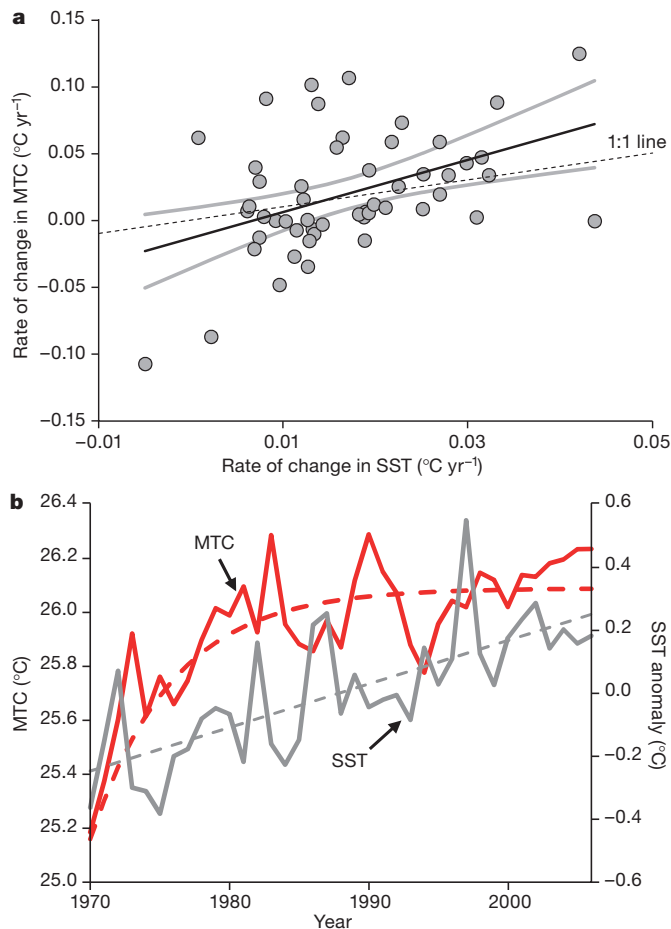


Figure 3 | Relationship between rates of change in MTC and SST between 1970 and 2006 in 52 LMEs. **a**, Rate of change of MTC was calculated from the ensemble GAMM in each LME (slope of linear regression, $P < 0.005$, $R^2 = 0.19$ (coefficient of determination)). The black line shows the mean and the grey lines delineate the 95% confidence interval. **b**, Changes in MTC (red) and SST anomalies (grey) in tropical LMEs. The dashed lines are fitted with asymptotic and linear models for the MTC and SST anomalies, respectively.

oceanographic variability and fishing effort, both of which are known to affect fisheries strongly^{21,22}. In addition, we repeated the analysis using a subset of 698 species with their temperature preference predicted from the above and an alternative approach to species distribution modelling (Supplementary Information).

Overall, the MTC in the 52 LMEs from the ensemble GAMM analysis increased at an average rate of 0.19°C per decade between 1970 and 2006 (Fig. 2), and the MTC in non-tropical LMEs increased at a rate of 0.23°C per decade. Specifically, MTC increased consistently in LMEs in the northeast Pacific Ocean (0.48°C per decade) and the northeast Atlantic Ocean (0.49°C per decade), where SST increased by 0.20 and 0.26°C per decade, respectively. For each LME, and on the basis of the results from the ensemble GAMMs, we ranked the relative importance of SST, fishing effort and large-scale oceanographic indices as factors accounting for the changes in MTC over time. The results suggested that there was no significant difference between these factors, in terms of the frequency distribution of LMEs with different rankings (Kruskal–Wallis rank-sum test, $P > 0.05$; Supplementary Fig. 1). However, there were regional differences in their importance (Supplementary Table 1).

MTC changes in the 52 LMEs were significantly related to the rate of SST changes between 1970 and 2006 ($P < 0.05$; Fig. 3 and Supplementary Table 2). The relationships remained significant when the best models (lowest Akaike information criteria) were selected instead of the model ensemble (Supplementary Table 2) and when the twenty-fifth and seventy-fifth percentiles of the temperature preference profiles

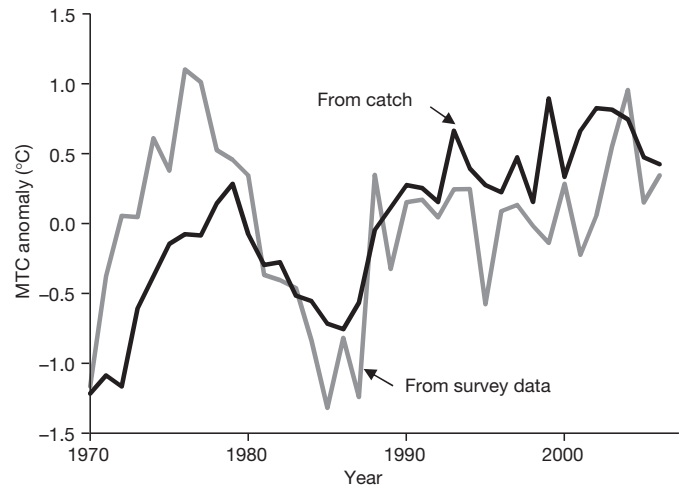


Figure 4 | Comparison of MTC calculated from fisheries catch data and relative abundance calculated from scientific survey data. MTC is expressed as an anomaly relative to the mean of the time series, and relative abundance is expressed relative to the average between 1970 and 2006. The type of data used had no effect on the rate of change in MTC (ANCOVA, $P > 0.1$).

of the species, instead of the median values, were used in calculating the MTC (Supplementary Fig. 2). Using the alternative species distribution modelling approach to calculate species' temperature preference has no significant effect on the relationship between changes in MTC and SST ($P > 0.1$; Methods and Supplementary Information).

Tropical LMEs showed, overall, an asymptotic pattern of MTC change, with a reduction in the proportion of subtropical species in catches (Fig. 3). Average MTC from 14 tropical LMEs increased rapidly between 1970 and 1980 by around 0.6°C , and subsequently stabilized at around 26°C (Fig. 3). Moreover, the average temperature tolerance range, calculated from the difference in temperature between the twenty-fifth and seventy-fifth percentiles of the temperature preference profile of each species (Supplementary Information), decreased significantly during the period. Subtropical species had wider temperature tolerances than did tropical species¹⁷ (Supplementary Fig. 3), providing further evidence to support the hypothesis that fisheries catches in the tropics are becoming inimical to subtropical species. Moreover, average SST change in the tropical LMEs increased consistently, at a rate of 0.14°C per decade.

A number of factors indicate that MTC is a valid proxy to examine changes in composition of catches in a region in relation to the temperature preference of the exploited animals. First, there was no difference (analysis of covariance (ANCOVA), $P > 0.1$; Supplementary Table 4) between using catch data and scientific bottom trawl survey data in the calculated rate of change in MTC in North Sea (Fig. 4 and Methods). Because the scientific survey data set can be viewed as a reliable indicator of relative abundance of animals in the ocean, our results therefore support the use of catch data to detect climate change signature in fisheries. However, because the availability of survey-based data was limited to a few well-studied regions²³, we were unable to use such data to analyse global fisheries. The relationship between the thermal preference of the species and the environmental temperature also corroborates the results of regional-scale studies using survey data^{14,24}.

Second, fishing efforts in many LMEs have been increasing continuously since the 1970s. This coincided with increases in SST, resulting in strong correlation between changes in SST and fishing effort in some LMEs. However, there is no evidence that fishing systematically alters MTC. Specifically, significant but weak relationships between maximum body size (positively related to vulnerability to fishing, in general) and the temperature preference of exploited species is found in only 19 LMEs, with the majority (13) of them showing a positive relationship, suggesting that the increasing MTC trend was not a result

of the depletion of large fish by fishing that was reported by many fisheries²⁵ (Supplementary Information).

Third, our findings are not an artefact of the quality of fisheries and environmental data, which vary between regions. In four of five areas in which misreporting of the level of catch was previously estimated, the estimated rates of change in MTC were not significantly different from those calculated from data sets without correction for misreporting. For the only area that showed a significant difference in rate of MTC change, MTCs calculated from both data sets show significant increase and the qualitative trend thus remains robust. Serious over-reporting of fisheries catches by China since the mid 1980s have been documented²⁶. Recalculating MTC in the South China Sea, the East China Sea and the Yellow Sea using catch data from 1970–1985 increased the goodness of fit of its significant relationship with SST. The use of different SST climatology to calculate species' temperature preference, moreover, had no significant effect on the findings (Supplementary Table 3).

Although fisheries catch statistics are usually not reported at stock level, an analysis using simulated species and stock distributions, temperature preferences and, subsequently, MTC changes showed that the existence of stock structure and stock-specific temperature preferences did not significantly bias the rate of change of MTC calculated from aggregated species-level data ($P > 0.1$; Supplementary Information). Potential phenotypic and evolutionary responses of species to warming, if occurring, would reduce the rate of change in MTC. The accuracy of the SST data set varied between regions as a result of different densities of measurements. Given these uncertainties, the fact that our analysis is able to detect a signature of climate change in global fisheries highlights the robustness of the underlying relationship.

Overall, our results suggest that change in the composition of marine fisheries catch is significantly related to temperature change in the ocean, with an increasing dominance of catches of warmer waters species at higher latitudes and a decrease in the proportion of catches of subtropical species in the tropics. Such changes in catch composition have direct implications for coastal fishing communities, particularly those in tropical developing countries, which tend to be socioeconomically vulnerable to the effects of climate change⁶. Continued warming in the tropics to a level that exceeds the thermal tolerance of tropical species may largely reduce catch potential in this region¹². This highlights the need to prioritize resources to develop adaptation plans immediately to minimize impacts on the economy and food security of tropical coastal communities.

METHODS SUMMARY

The mapped fisheries landings data were sourced from the Sea Around Us project²⁷, which utilized a range of sources including the Food and Agriculture Organization's (FAO) fisheries database supplemented by regional data sets (Supplementary Information). We inferred the thermal preference of each species on the basis of its modelled distribution^{28,29} (Supplementary Table 5 and Supplementary Information). The MTC was computed from the average inferred temperature preference of 990 species of exploited fishes and invertebrates weighted by their annual catch:

$$MTC_{yr} = \frac{\sum_i^n T_i C_{i,yr}}{\sum_i^n C_{i,yr}} \quad (1)$$

Here $C_{i,yr}$ is catch of species i in a specific region in year yr , T_i is the median temperature preference of species i and n is the total number of species. Similarly, MTC was calculated using survey data in the North Sea from equation (1) with $C_{i,yr}$ replaced by the relative abundance for species i in year yr . Relative abundance data were obtained from the ICES International Bottom Trawl Survey. A total of 55 species common to both the catch database and the survey data set were included in this analysis. We tested the effect of the use of catch or survey data in estimating changes in MTC over time using ANCOVA, accounting for temporal autocorrelation.

We used the GAMM (R package MuMIn) with a full model, $MTC = s(\text{Effort}) + s(\text{OI}) + \text{Year}$, where s is a spline smoothing function (Supplementary Information) with a temporal autocorrelation term (R function corAR1), allowing for a lag of up to three years in the response of MTC to these two variables and the time effect expressed by Year. To estimate the average rate of change in MTC from the coefficients of this time effect, we did not include a spline smoothing function for Year. Fishing effort data were sourced from the Sea Around Us project³⁰ (Supplementary Information).

Full Methods and any associated references are available in the online version of the paper.

Received 13 October 2012; accepted 5 April 2013.

- Cheung, W. W. L. *et al.* Climate change induced tropicalization of marine communities in Western Australia. *Mar. Freshw. Res.* **63**, 415–427 (2012).
- Wernberg, T. *et al.* An extreme climatic event alters marine ecosystem structure in a global biodiversity hotspot. *Nature Clim. Change* **3**, 78–82 (2013).
- Perry, A. L., Low, P. J., Ellis, J. R. & Reynolds, J. D. Climate change and distribution shifts in marine fishes. *Science* **308**, 1912–1915 (2005).
- Dulvy, N. K. *et al.* Climate change and deepening of the North Sea fish assemblage: a biotic indicator of warming seas. *J. Appl. Ecol.* **45**, 1029–1039 (2008).
- Belkin, I. M. Rapid warming of large marine ecosystems. *Prog. Oceanogr.* **81**, 207–213 (2009).
- Allison, E. H. *et al.* Vulnerability of national economies to the impacts of climate change on fisheries. *Fish Fish.* **10**, 173–196 (2009).
- Sumaila, U. R., Cheung, W. W. L., Lam, V. W. Y., Pauly, D. & Herrick, S. Climate change impacts on the biophysics and economics of world fisheries. *Nature Clim. Change* **1**, 449–456 (2011).
- Simpson, S. D. *et al.* Continental shelf-wide response of a fish assemblage to rapid warming of the sea. *Curr. Biol.* **21**, 1565–1570 (2011).
- Edwards, M. & Richardson, A. J. Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature* **430**, 881–884 (2004).
- Cheung, W. W. L. *et al.* Shrinking of fishes exacerbates impacts of global ocean changes on marine ecosystems. *Nature Clim. Change* **3**, 254–258 (2013).
- Harley, C. D. G. Climate change, keystone predation, and biodiversity loss. *Science* **334**, 1124–1127 (2011).
- Cheung, W. W. L. *et al.* Large-scale redistribution of maximum fisheries catch potential in the global ocean under climate change. *Glob. Change Biol.* **16**, 24–35 (2010).
- Blanchard, J. *et al.* Potential consequences of climate change for primary production and fish production in large marine ecosystems. *Phil. Trans. R. Soc. B* **367**, 2979–2989 (2012).
- Pinsky, M. & Fogarty, M. Lagged social-ecological responses to climate and range shifts in fisheries. *Clim. Change* **115**, 883–891 (2012).
- Cheung, W. W. L., Pinnegar, J., Merino, G., Jones, M. C. & Barange, M. Review of climate change impacts on marine fisheries in the UK and Ireland. *Aquat. Conserv. Mar. Freshwat. Ecosyst.* **22**, 368–388 (2012).
- Sunday, J. M., Bates, A. E. & Dulvy, N. K. Global analysis of thermal tolerance and latitude in ectotherms. *Proc. R. Soc. Lond. B* **278**, 1823–1830 (2011).
- Pörtner, H. O. & Farrell, A. P. Physiology and climate change. *Science* **322**, 690–692 (2008).
- Pörtner, H. O. & Knust, R. Climate change affects marine fishes through the oxygen limitation of thermal tolerance. *Science* **315**, 95–97 (2007).
- Ben Rais Lasram, F. *et al.* The Mediterranean Sea as a 'cul-de-sac' for endemic fishes facing climate change. *Glob. Change Biol.* **16**, 3233–3245 (2010).
- Collie, J. S., Wood, A. D. & Jeffries, H. P. Long-term shifts in the species composition of a coastal fish community. *Can. J. Fish. Aquat. Sci.* **65**, 1352–1365 (2008).
- McGowan, J. A., Cayan, D. R. & Dorman, L. M. Climate-ocean variability and ecosystem response in the northeast Pacific. *Science* **281**, 210–217 (1998).
- Pauly, D. *et al.* Towards sustainability in world fisheries. *Nature* **418**, 689–695 (2002).
- Costello, C. *et al.* Status and solutions for the world's unassessed fisheries. *Science*, (2012).
- Howell, P. & Auster, P. J. Phase shift in an estuarine finfish community associated with warming temperatures. *Mar. Coast. Fish.* **4**, 481–495 (2012).
- Cheung, W. W. L., Watson, R., Morato, T., Pitcher, T. J. & Pauly, D. Intrinsic vulnerability in the global fish catch. *Mar. Ecol. Prog. Ser.* **333**, 1–12 (2007).
- Watson, R. & Pauly, D. Systematic distortions in world fisheries catch trends. *Nature* **414**, 534–536 (2001).
- Watson, R., Kitchingman, A., Gelchu, A. & Pauly, D. Mapping global fisheries: sharpening our focus. *Fish Fish.* **5**, 168–177 (2004).
- Jones, M., Dye, S., Pinnegar, J., Warren, R. & Cheung, W. W. L. Modelling commercial fish distributions: prediction and assessment using different approaches. *Ecol. Modell.* **225**, 133–145 (2012).
- Röckmann, C., Dickey-Collas, M., Payne, M. R. & van Hal, R. Realized habitats of early-stage North Sea herring: looking for signals of environmental change. *ICES J. Mar. Sci.* **68**, 537–546 (2011).
- Watson, R. A. *et al.* Global marine yield halved as fishing intensity redoubles. *Fish Fish.* (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements W.W.L.C. acknowledges funding support from the National Geographic Society and the Natural Sciences and Engineering Research Council of Canada. R.W. and D.P. were supported by the Pew Charitable Trust through the Sea Around Us project. We are grateful to S. Pauly and D. Palomares for reviewing the manuscript and providing the Aquamap distributions from FishBase, respectively.

Author Contributions W.W.L.C. and D.P. designed the study. W.W.L.C. conducted the analysis. R.W. and D.P. provided the fisheries catch and effort data from the Sea Around Us project. All authors contributed to the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.W.L.C. (w.cheung@fisheries.ubc.ca).

METHODS

Global catch and effort data. We use the terms fisheries catch and fisheries landings interchangeably, but strictly these both refer here to landings; the weight of marine fishes and invertebrates that were caught and retained. The mapped fisheries landings and effort data were sourced from the Sea Around Us project²⁷ (Supplementary Information). Landings data were not used to estimate temperature preference of the species, and temperature data were not used to infer species distributions. Effort data were standardized and collated on the basis of fishing boat engine power (watts) and fishing days³⁰ (Supplementary Information).

Inferring species' thermal preference. We inferred the thermal preference of each species from its modelled distribution (Supplementary Table 5). First, we modelled the present (1970–2000) distribution of each species using an algorithm developed by the Sea Around Us project and documented in ref. 28. The algorithm estimated the relative abundance (on a $30' \times 30'$ latitude–longitude grid) of a species in each spatial cell. Input parameters for each species considered in the model included the species' maximum and minimum depth limits, northern and southern latitudinal range limits, an index of association to major habitat types (seamounts, estuaries, inshore, offshore, continental shelf, continental slope and the abyssal) and known occurrence boundaries. For pelagic species, seasonal (summer and winter) distributions were considered. The parameter values of each species were derived from data in online databases, mainly FishBase (<http://www.fishbase.org>) and SeaLifeBase (<http://www.sealifebase.org>). Catch and temperature data were not used to model the current species distributions. Each modelled species distribution was normalized and overlaid over the SST climatology from the Hadley Centre SST data set for 1970–2000²⁹. The predicted species distribution is considered to be representative of the distribution of relative abundance. The temperature preference profile at SST bin i (p_i) of each species was calculated from the total relative abundance, K_i , and range area, A_i , at SST bin i (Supplementary Information): $p_i = (K_i/A_i) / \sum_i (K_i/A_i)$.

The median and the twenty-fifth and seventy-fifth percentiles of the temperature preference were calculated. The modelled species distribution and temperature preference represents an average of the species' life stages. The analysis was repeated with alternative estimate climatologies for 1951–1960 and 1991–2000. The choice of climatology has no significant effect (ANCOVA, $P > 0.1$) on the relationship between rate of change in MTC and rate of change in SST (Supplementary Table 3). The analysis is also repeated using an alternative approach to species distribution modelling (Supplementary Information).

Calculation of MTC. The MTC was computed from the average inferred temperature preference of 990 species of exploited fishes and invertebrates weighted by their annual catch:

$$MTC_{yr} = \frac{\sum_i^n T_i C_{i,yr}}{\sum_i^n C_{i,yr}} \quad (1)$$

Here $C_{i,yr}$ is catch of species i in a specific region in year yr , T_i is the median temperature preference of species i and n is the total number of species.

In the North Sea LME, MTC was calculated from equation (1) with $C_{i,yr}$ replaced by the relative abundance for species i in year yr . Relative abundance data were obtained from the ICES International Bottom Trawl Survey. A total of 55 species that coexist in both the catch database and the survey data set were included in this analysis. We tested whether the changes in MTC calculated from catch data and scientific survey data were different, using ANCOVA and accounting for temporal autocorrelation.

Large-scale oceanographic indices. Indices representing large-scale oceanographic conditions for each of the six ocean basins were included in this study (Supplementary Information). For LMEs in the Pacific Ocean, their relationship with the Pacific decadal oscillation was used. For LMEs in the North Atlantic and South Atlantic, the North Atlantic oscillation index and, respectively, the dipole index were used. For LMEs in the Indian Ocean, the Indian Ocean dipole index was used. For LMEs in the Arctic, the summer sea-ice extent was used. For LMEs in the Southern Ocean, the Antarctic Oscillation index was used.

Accounting for the effects of fishing and large-scale oceanographic indices.

For changes in MTC in each LME, we accounted for the effect of changes in total fishing effort (E) and the corresponding large-scale oceanographic index (OI). We then used a GAMM (R package MuMIn) with a full model, $MTC = s(Effort + s(OI) + Year)$, where s represents a spline smoothing function (Supplementary Information), allowing for a lag of up to three years in the response of MTC to these two variables and the time effect Year, and including an autoregressive term to account for temporal autocorrelation. To estimate the average rate of change in MTC from the coefficients of the time effect, we did not include a spline smoothing function for the variable Year. These models were analysed using a multi-model analysis framework (R package MuMIn). There were significant autocorrelations of 1-yr lag in the MTC time series for all LMEs ($P < 0.05$). Rates of MTC changes (coefficient of Year) were estimated using a multi-model ensemble mean with weighting (W) calculated from $W_i = \exp[(AIC_{\min} - AIC_i)/2]$, where AIC_{\min} is the best model (that with minimum Akaike information criterion) and AIC_i was based on alternative model i . In addition, the rates of MTC change were also estimated from the model with the lowest Akaike information criterion. The relationship between the rates of change in MTC and SST remained significant with these alternative estimates.

Protective astrogenesis from the SVZ niche after injury is controlled by Notch modulator Thbs4

Eric J. Benner¹, Dominic Luciano^{2,3}, Rebecca Jo^{1,2}, Khadar Abdi², Patricia Paez-Gonzalez², Huaxin Sheng⁴, David S. Warner⁴, Chunlei Liu^{5,6}, Cagla Eroglu^{2,3,7} & Chay T. Kuo^{1,2,3,7,8}

Postnatal/adult neural stem cells (NSCs) within the rodent subventricular zone (SVZ; also called subependymal zone) generate doublecortin (Dcx)⁺ neuroblasts that migrate and integrate into olfactory bulb circuitry^{1,2}. Continuous production of neuroblasts is controlled by the SVZ microenvironmental niche^{3,4}. It is generally thought that enhancing the neurogenic activities of endogenous NSCs may provide needed therapeutic options for disease states and after brain injury. However, SVZ NSCs can also differentiate into astrocytes. It remains unclear whether there are conditions that favour astrogenesis over neurogenesis in the SVZ niche, and whether astrocytes produced there have different properties compared with astrocytes produced elsewhere in the brain⁵. Here we show in mice that SVZ-generated astrocytes express high levels of thrombospondin 4 (Thbs4)^{6,7}, a secreted homopentameric glycoprotein, in contrast to cortical astrocytes, which express low levels of Thbs4. We found that localized photothrombotic/ischaemic cortical injury initiates a marked increase in Thbs4^{hi} astrocyte production from the postnatal SVZ niche. Tamoxifen-inducible *nestin-creERtm4* lineage tracing demonstrated that it is these SVZ-generated Thbs4^{hi} astrocytes, and not Dcx⁺ neuroblasts, that home-in on the injured cortex. This robust post-injury astrogenic response required SVZ Notch activation modulated by Thbs4 via direct Notch1 receptor binding and endocytosis to activate downstream signals, including increased Nfia transcription factor expression important for glia production⁸. Consequently, Thbs4 homozygous knockout mice (Thbs4^{KO/KO}) showed severe defects in cortical-injury-induced SVZ astrogenesis, instead producing cells expressing Dcx migrating from SVZ to the injury sites. These alterations in cellular responses resulted in abnormal glial scar formation after injury, and significantly increased microvascular haemorrhage into the brain parenchyma of Thbs4^{KO/KO} mice. Taken together, these findings have important implications for post-injury applications of endogenous and transplanted NSCs in the therapeutic setting, as well as disease states where Thbs family members have important roles^{9,10}.

We want to understand neurogenesis/astrogenesis choice in the postnatal SVZ niche, and the cellular properties of SVZ-generated astrocytes. CNS astrocytes secrete thrombospondins (Thbs), a family of homotrimeric and homopentameric proteins⁷: we asked whether SVZ-generated astrocytes express different Thbs protein(s) than their cortical counterparts. We started by establishing primary SVZ adherent cultures from *nestin-creERtm4* (N4); *rosa26reporter-tdTomato* (R26R-tdTomato) mice, after postnatal day 7 (P7) tamoxifen injection. We showed previously that this N4 line can inducibly and efficiently lineage-trace the progeny of postnatal/adult SVZ NSCs^{4,11} (Supplementary Fig. 1 and Supplementary Video 1). After 5 days of differentiation, tdTomato⁺ lineage-traced GFAP⁺ astrocytes showed strong immunohistochemical (IHC) staining for Thbs4 (Supplementary Fig. 2a). To

confirm the specificity of this antibody staining, we performed the same culture experiment using Thbs4^{KO/KO} mutant mice¹², which revealed no Thbs4 immunofluorescence under identical experimental/imaging conditions (Supplementary Fig. 2b). Western blot analyses confirmed these findings (Supplementary Fig. 2c). Comparison of differentiated SVZ astrocyte cultures to primary astrocytes collected from the cortex showed selective Thbs4 expression by SVZ astrocytes (Fig. 1a). This difference in Thbs4 expression was further demonstrated by quantitative polymerase chain reaction (qPCR) analyses of FACS-purified GFP⁺ cortical versus SVZ astrocytes from GFAP-GFP transgenic mice (Fig. 1b). IHC staining on P14 brain sections indicated that Thbs4, although co-labelling with SVZ GFAP⁺ astrocytes, did not co-localize with Mash1⁺ transiently amplifying progenitors, Dcx⁺ neuroblasts, nor NG2⁺ or Olig2⁺ populations in the SVZ niche (Supplementary Fig. 3a–e).

We next tested whether SVZ-generated GFAP⁺ astrocytes are fated to express Thbs4 *in vivo*. We transplanted second passage SVZ NSC

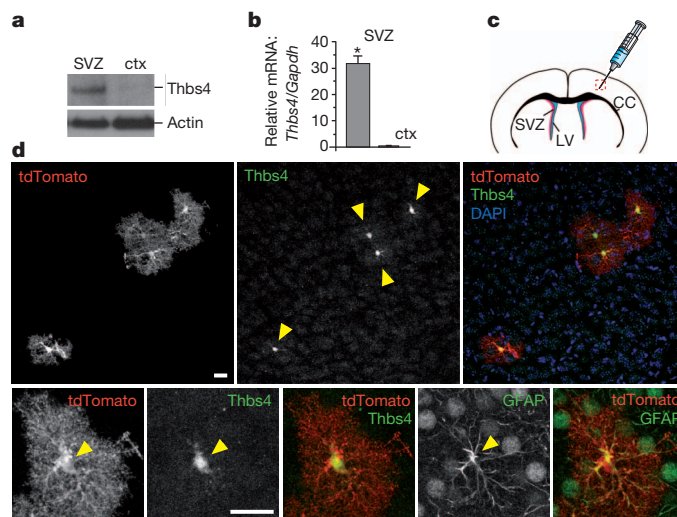


Figure 1 | SVZ generation of Thbs4^{hi} astrocytes. **a**, Western blot analysis of Thbs4 protein levels in differentiated primary SVZ and cortical (ctx) astrocyte cultures. **b**, qPCR analyses of Thbs4 levels in FACS-sorted SVZ versus cortical GFP⁺ astrocytes from GFAP-GFP transgenic mice. **P* < 0.001, *n* = 5, Student's *t*-test; error bars indicate s.e.m. **c**, Schematic representation of cortical transplantation strategy. CC, corpus callosum; LV, lateral ventricle. **d**, Thbs4, tdTomato and GFAP IHC antibody staining of brain sections—2–4 weeks after mice were transplanted with lineage-traced primary SVZ NSC cultures derived from tamoxifen-induced *nestin-creERtm4*; R26R-tdTomato (N4; R26R-tdTomato) mice—showing co-localization between tdTomato, Thbs4 and GFAP (arrowheads). Scale bar: 20 μ m.

¹George and Jean Brumley Neonatal-Perinatal Research Institute, Department of Pediatrics, Duke University School of Medicine, Durham, North Carolina 27710, USA. ²Department of Cell Biology, Duke University School of Medicine, Durham, North Carolina 27710, USA. ³Department of Neurobiology, Duke University School of Medicine, Durham, North Carolina 27710, USA. ⁴Department of Anesthesiology, Duke University School of Medicine, Durham, North Carolina 27710, USA. ⁵Brain Imaging and Analysis Center, Department of Radiology, Duke University School of Medicine, Durham, North Carolina 27710, USA. ⁶Department of Radiology, Duke University School of Medicine, Durham, North Carolina 27710, USA. ⁷Duke Institute for Brain Sciences, Duke University School of Medicine, Durham, North Carolina 27710, USA. ⁸Preston Robert Tisch Brain Tumor Center, Duke University School of Medicine, Durham, North Carolina 27710, USA.

cultures collected from tamoxifen-induced *N4; R26R-tdTomato* mice into the cerebral cortex of wild-type/non-transgenic mice (Fig. 1c, d). Unlike transplantations into the SVZ, which generated neuroblasts that migrate to the olfactory bulb⁴ (Supplementary Fig. 4), *tdTomato*⁺ cells transplanted into the cortex after 2–4 weeks gave rise to GFAP⁺Thbs4^{hi} astrocytes, contrasting with resident cortical astrocytes which were Thbs4^{low} (Fig. 1d). Of the 402 *N4* lineage-traced *tdTomato*⁺ cells counted over multiple experiments, 96.8% (389) showed GFAP expression. Strong Thbs4⁺ IHC co-staining was detected in 98.2% (382) of these GFAP⁺ cells.

Next, we investigated under which conditions the SVZ niche favoured Thbs4^{hi} astrocyte production. SVZ NSCs are thought to respond to middle cerebral artery occlusion-induced stroke by producing Dcx⁺ neuroblasts that migrate into the striatum¹³; however, some studies have also reported gliogenic responses after injury^{14,15}. To generate precisely localized cortical injuries directly over the SVZ niche we used photothrombosis, a well-defined ischaemic injury model¹⁶ (Fig. 2a). We performed these injuries using *N4; R26R-tdTomato* mice after P6 tamoxifen injection: this timing allowed us specifically to target/lineage-trace SVZ NSCs, while minimizing Cre-reporter labelling of cortical cells¹¹. We induced cortical injuries at P12 and examined the SVZ responses at 3 and 14 days post-injury (d.p.i.), using 3,3'-diaminobenzidine (DAB) IHC staining against *tdTomato*, allowing simultaneous visualization of *tdTomato*-reporter expression and brain tissue histology. At 3 d.p.i. we did not see significant

morphological changes to lineage-traced *tdTomato*⁺ cells around the SVZ (Fig. 2b). However, by 14 d.p.i. we observed robust populations of *tdTomato*⁺ cells between SVZ and the cortical injury site (Fig. 2b). This delayed reaction was specific to SVZ/hemisphere ipsilateral to the injury, as we did not detect obvious changes to *tdTomato*⁺ cells in the contralateral cortex, nor did we see changes in the ipsilateral cortex in sham-treated control brains (Supplementary Fig. 5a, b). As a further control, we did not detect *tdTomato*⁺ cells in the injured cortex of *N4; R26R-tdTomato* mice without tamoxifen injection (Supplementary Fig. 5c). We also did not observe *tdTomato*⁺ cells next to superficial cortical injuries, suggesting a correlation between injury severity and SVZ responses (data not shown). All analyses were performed on cortical injuries that did not breach the corpus callosum (determined by sectioning through the entire injured area), limiting potential spill-over of SVZ niche neuroblasts into the injured areas.

IHC staining showed that most of the *tdTomato*⁺ cells next to the injury site 14 d.p.i. had complex morphologies, labelled strongly with anti-Thbs4 antibody, and co-stained with GFAP (Fig. 2c, d). Although we did not observe *tdTomato*⁺ cells at the injury site 3 d.p.i., western blotting and IHC staining of SVZ tissue showed that, at this early time point, the ipsilateral SVZ had upregulated Thbs4 expression as compared to the contralateral side (Fig. 2e, f). This Thbs4 expression remained localized to GFAP⁺ astrocytes (Supplementary Fig. 3f). Repeating these injury experiments using the *GFAP-cre* driver (JAX no. 004600) instead of *N4*, we observed similarly robust *tdTomato*⁺ lineage-traced cells around the injury site 14 d.p.i. (Supplementary Fig. 6), although they probably represent a heterogeneous cell population, as the *GFAP-cre* driver also efficiently lineage-traced cortical cells (Supplementary Fig. 6). To investigate whether cortical astrocytes can upregulate Thbs4 expression to levels seen in SVZ astrocytes after injury, we performed qPCR analyses 3 d.p.i. on FACS-sorted GFP⁺ cortical versus SVZ astrocytes from *GFAP-GFP* mice (before significant migration from SVZ), which showed that SVZ astrocytes remained the Thbs4^{hi} population (Supplementary Fig. 7).

We next investigated the molecular signals responsible for controlling this injury-induced Thbs4^{hi} astrogenesis from the SVZ niche. The Notch pathway is an important regulator of SVZ NSC and niche function^{11,17}, as well as neurogenesis versus gliogenesis choice¹⁸. To determine whether Notch activation is required for Thbs4^{hi} astrocyte production after cortical injury, we inducibly deleted *RBPjk*¹⁹, the Notch intracellular domain (NICD) co-transcriptional activator, using the *N4* transgene. We crossed *N4; RBPjk*^{KO/+} mice to *R26R-tdTomato*^{fl/fl}; *RBPjk*^{fl/fl} mice and treated new mouse litters with tamoxifen at P6. We then compared lineage-traced *tdTomato*⁺ cells around the cortical injury site between *N4; R26R-tdTomato*; *RBPjk*^{fl/fl} (control) and *N4; R26R-tdTomato*; *RBPjk*^{KO/fl} inducible mutant (iKO) mice at 14 d.p.i. DAB staining revealed that although there were robust *tdTomato*⁺ cells between the SVZ and cortical injury site in the controls (Fig. 3a), this response was blunted in *RBPjk* iKO mice (Fig. 3a, b). We next asked if increased Notch signalling can result in enhancement of SVZ response after injury. Using *N4; R26R-tdTomato*; *R26R-NICD*^{fl/+} mice to over-express NICD in lineage-traced SVZ cells after tamoxifen injection, we repeated the cortical injury experiments. DAB staining 14 d.p.i. showed significantly increased numbers of lineage-traced *tdTomato*⁺ cells at the injury site in NICD-overexpressing mice (Fig. 3a, b). Consistent with these results, IHC staining of wild-type brain sections after cortical photothrombosis showed upregulated Notch1 receptor expression in the ipsilateral SVZ 2 d.p.i. (Supplementary Fig. 8). Furthermore, western blot analyses 3 d.p.i. revealed an increase in NICD protein in the ipsilateral SVZ (Fig. 3c).

Ependymal cells, labelled/lineage-traced by *FOXJ1*-based reporters/drivers, have been reported to respond to ischaemic injury by generating progeny that migrate to the injury site²⁰. As this *N4* line also targets ependymal cells^{4,11}, thus potentially lineage-tracing ependymal progeny after injury, we performed identical cortical injuries in *FOXJ1-creER*²²; *R26R-tdTomato* mice after P6 tamoxifen injection. In repeated

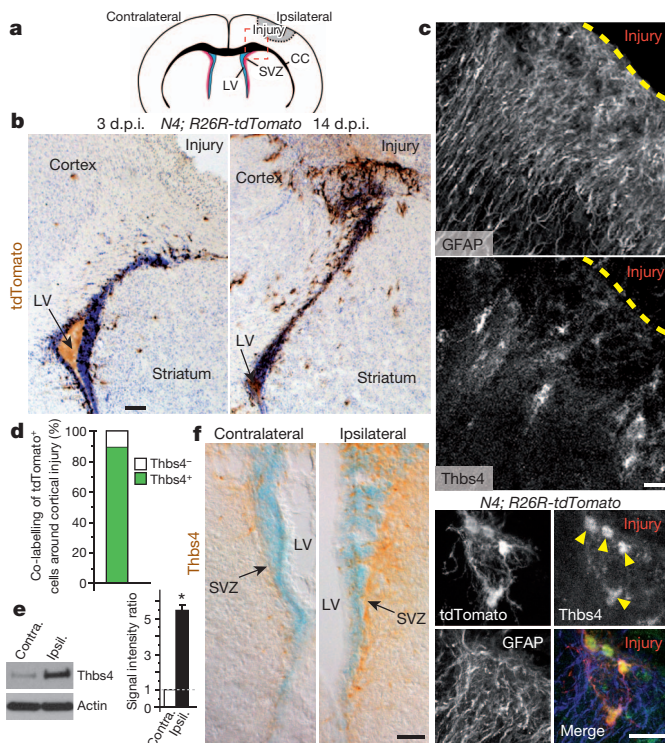


Figure 2 | Thbs4^{hi} astrocyte production after photothrombotic cortical injury. **a**, Schematic representation of photothrombosis injury model, with dashed box indicating region of imaging in **b**. **b**, DAB IHC staining for *tdTomato* in *N4; R26R-tdTomato* mice induced with tamoxifen, showing a delayed activation of lineage-traced *tdTomato*⁺ cells to injury site 14 d.p.i. **c**, Coronal sections of injury site 14 d.p.i., IHC stained for GFAP, Thbs4 and *tdTomato*, showing that lineage-traced *tdTomato*⁺ cells adjacent to injury are Thbs4^{hi}GFAP⁺ astrocytes (arrowheads). **d**, Quantitative analyses of total *tdTomato*⁺ cells at injury site expressing Thbs4 14 d.p.i. (88.10 ± 1.99% s.d., *n* = 3 mice). **e**, Western blot and quantitative analyses of Thbs4 protein levels in SVZ tissues 3 d.p.i. **P* < 0.001, *n* = 5, Student's *t*-test, error bars indicate s.e.m. **f**, DAB IHC staining for Thbs4 expression 3 d.p.i., Nissl-counterstained. Ipsilateral and contralateral SVZ from the same brain section, imaged under identical conditions. Scale bars: 100 μm (**b**); 20 μm (**c**); 50 μm (**f**).

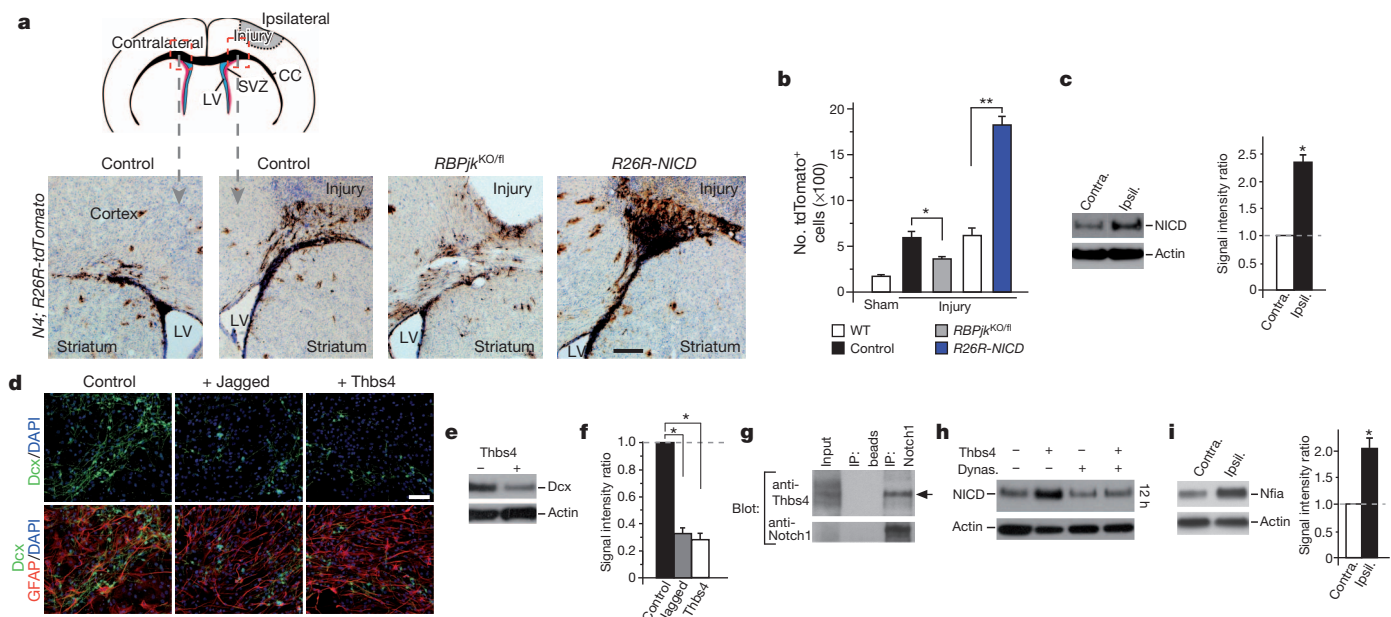


Figure 3 | Notch signalling and regulation of injury-induced SVZ astrogenesis. **a**, Photothrombotic cortical injury model, with areas of imaging indicated by dashed boxes. Representative sections 14 d.p.i. showing DAB IHC staining for tdTomato from tamoxifen-induced *N4*; *R26R-tdTomato*; *RBPjk^{KO/RI}* (control); *N4*; *R26R-tdTomato*; *RBPjk^{KO/RI}* (*RBPjk^{KO/RI}*); and *N4*; *R26R-tdTomato*; *R26R-NICD* mice. In control panels, ipsilateral and contralateral hemispheres are shown, corresponding to boxes in diagram (indicated by dashed arrows). **b**, Quantification of tdTomato⁺ cells above the corpus callosum in each genetic background 14 d.p.i. **P* < 0.05, *n* = 6; ***P* < 0.001 (wild type (WT), *n* = 8; NICD, *n* = 5), Student's *t*-test; error bars indicate s.e.m. **c**, Western blot analyses of Notch intracellular domain (NICD) protein levels in SVZ tissues collected 3 d.p.i., showing upregulation in the ipsilateral over the contralateral side from the same brain. **P* < 0.005, *n* = 5, Student's *t*-test, error bars indicate s.e.m. **d**, Differentiation of SVZ adherent neural stem cell cultures, with or without Jag1-Fc, and/or Thbs4 added. **e**, Western blot analyses comparing Dcx levels after 5 days of culture differentiation. **f**, Quantification of Dcx levels on western blots. **P* < 0.001, *n* = 5, Student's *t*-test, error bars indicate s.e.m. **g**, Freshly isolated SVZ tissue 3 d.p.i.: immunoprecipitation (IP) with control beads or anti-Notch1 antibody, and blotted with anti-Thbs4 or anti-Notch1 antibodies, detecting Thbs4 pull down (arrow). **h**, Thbs4 induction of NICD during *in vitro* differentiation with or without dynasore, 12 h after stimulation. **i**, Western blot analyses of Nfia levels in SVZ tissues collected 3 d.p.i., showing upregulation in the ipsilateral over the contralateral side from the same brain. **P* < 0.005, *n* = 5, Student's *t*-test, error bars indicate s.e.m. Scale bars: 200 μm (a); 50 μm (d).

experiments we did not observe migration of tdTomato⁺ populations from SVZ to the cortical injury sites either early (3 d.p.i.) or late (14 d.p.i.) (Supplementary Fig. 9a and data not shown). The *FOXJ1-creER¹²* driver also labelled mature cortical neurons, but this expression was independent from injury-induced cell proliferation (Supplementary Fig. 9a and data not shown). To rule out the possibility that modification of Notch signalling allowed ependymal cells to generate progeny responding to injury, we performed cortical injury experiments in *RBPjk* iKO and *R26R-NICD* overexpression backgrounds using the *FOXJ1-creER¹²* driver. In both genetic backgrounds we did not observe migration of tdTomato⁺ cells from SVZ to the injury sites (Supplementary Fig. 9b).

Although thrombospondins, to our knowledge, have not been previously linked to Notch function *in vivo*, Thbs2 has been shown to interact physically with the Notch3 receptor *in vitro*²¹. Thus, we wondered whether Thbs4 has a direct role in modulating Notch activity after cortical injury. To test this, we first differentiated early passage SVZ NSC cultures in the presence of immobilized Jagged1 (Jag1)-Fc. Because transient Notch activation is known to decrease neurogenesis¹⁸, immobilized Jag1-Fc resulted in decreased neuroblast production compared to untreated control cultures (Fig. 3d, f). We repeated this assay with the addition of recombinant Thbs4 alone in the medium, and observed a similarly robust decrease in neuroblast production (Fig. 3d, f). This was confirmed by western blot analyses for Dcx protein after 5 days of differentiation (Fig. 3e, f). DAPT or DBZ, common Notch pathway inhibitors, conversely increased production of neuroblasts from differentiating SVZ cultures (Supplementary Fig. 10 and data not shown). Thbs4 was unable to blunt DAPT- or DBZ-mediated effects (Supplementary Fig. 10), indicating that the function of Thbs4 in this setting required Notch activity.

Because SVZ NSCs express the Notch1 receptor¹⁷, we asked whether Thbs4 and Notch1 can physically interact. Using varied concentrations of purified Thbs4 protein and recombinant Notch1-Fc fragment (including the first 13 extracellular EGF-like repeats of Notch1), pull-down experiments showed direct Thbs4-Notch1-Fc interactions *in vitro* (Supplementary Fig. 11a). mIgG2a, an isotype-specific control for Notch1-Fc, as well as purified Thbs2 protein did not show Notch1-Fc interactions (Supplementary Fig. 11a). We next collected cellular extracts from primary SVZ NSC cultures with or without addition of Thbs4 in the medium. Immunoprecipitation with anti-Notch1 antibody was able to pull down Thbs4 from the culture lysates in the +Thbs4 condition (Supplementary Fig. 11b). Whereas under physiological conditions immunoprecipitation from freshly isolated SVZ tissues using anti-Notch1 antibody did not detect significant Thbs4 pull down (Supplementary Fig. 11c), similar immunoprecipitation experiments 3 d.p.i. revealed Notch1-Thbs4 interactions (Fig. 3g). Efficient Notch signalling requires receptor endocytosis into endosomes²², and Thbs4 as a pentameric protein may promote Notch endocytosis and activation. To test this possibility, we cultured primary SVZ NSCs in the presence of both Thbs4 and late endocytosis blocker dynasore. Addition of Thbs4 alone resulted in NICD upregulation, as detected by western blotting 15 min after incubation, and lasting for several hours thereafter (Fig. 3h and Supplementary Fig. 11d). This NICD increase was blocked by dynasore (Fig. 3h). By 36 h after stimulation, Thbs4-induced NICD upregulation had subsided, and as expected with dynasore added there was low NICD protein level with or without Thbs4 in the medium (Supplementary Fig. 11e). The transcription factor Nfia is an important regulator of gliogenesis downstream of Notch signalling^{8,23}. After Thbs4 addition during SVZ NSC differentiation *in vitro*, we detected increased Nfia

expression over control conditions (Supplementary Fig. 11f). Similarly, western blotting of SVZ tissues 3 d.p.i. showed increased Nfia protein levels in the ipsilateral compared with contralateral SVZ (Fig. 3i), consistent with injury-induced SVZ astrogenic response (Fig. 2e, f).

Our findings raised the possibility that *Thbs4*, after cortical injury, is itself necessary for Notch-mediated SVZ astrogenesis. To test this, we first performed western blot analyses on SVZ tissue from *Thbs4*^{KO/KO} mice 3 d.p.i. Although under physiological conditions we did not detect noticeable differences in SVZ NICD expression levels between *Thbs4*^{KO/+} (control) and *Thbs4*^{KO/KO} mutant mice (Supplementary Fig. 12a), these mutants lacked increase in NICD expression in the ipsilateral SVZ 3 d.p.i. (Fig. 4a) unlike wild-type mice (Fig. 3c). Furthermore, *Thbs4*^{KO/KO} mice did not show an appreciable increase in Nfia expression in the ipsilateral SVZ 3 d.p.i. (Fig. 4a). We next tamoxifen-induced *N4; R26R-tdTomato; Thbs4*^{KO/KO} mice to lineage-trace responding SVZ-derived cells after cortical injury, according to the same protocol described above. At 14 d.p.i., the tdTomato⁺ cells around the injury site showed a surprising neuroblast-like morphology, and mostly co-labelled with Dcx rather than GFAP (Fig. 4b, c and Supplementary Fig. 12b). This is in marked contrast to cortical injuries in control mice, where few Dcx⁺ cells were detected next to the injury site (SVZ neuroblast staining was used as an internal control for Dcx antibody; Fig. 4c and Supplementary Fig. 13). The observed mixture of tdTomato⁻ and tdTomato⁺ Dcx⁺ cells at the cortical injury site is consistent with the clonal nature of *N4*-mediated recombination after tamoxifen induction^{4,11}.

We wanted to understand whether this change from SVZ astrogenesis to neurogenesis in *Thbs4*^{KO/KO} mice had significant effects on injury-induced sequelae. In preparing the injured brains from *Thbs4*^{KO/KO} mice, we noticed significant haemorrhage around the injured cortical areas. We next performed brain magnetic resonance imaging (MRI) analyses on control and *Thbs4*^{KO/KO} littermates. Using three-dimensional structural MRI on brains inside intact skulls, we can identify the location and extent of cortical injuries (Supplementary Fig. 14a, b). In isotropic diffusion-weighted images, the boundaries of lesions were hyperintense in both control and *Thbs4*^{KO/KO} mice (Supplementary Fig. 14a, b), demonstrating classical characteristics of stroke infarct²⁴. The interiors of these lesions appeared hypointense as result of increased oedema, and quantitative maps of diffusion coefficients confirmed hindered molecular diffusion in the infarcted areas but faster diffusion in the oedematous areas (Supplementary Fig. 14a, b). We did not detect significant mean diffusivity measurement differences between littermate control and *Thbs4*^{KO/KO} mice in the infarcted or oedematous cortical areas, either 2 d.p.i. (data not shown) or 8 d.p.i. (Fig. 4e and Supplementary Fig. 14d).

Iron in blood deoxyhaemoglobin is paramagnetic, thus increasing tissue magnetic susceptibility in the presence of haemorrhage²⁵. Using three-dimensional spoiled-gradient-recalled (SPGR) sequence to measure magnetic properties, the oedematous areas of *Thbs4*^{KO/KO} mice showed both faster image-intensity decay and faster *R*₂^{*} relaxation rate (rate of decay) as compared to littermate controls (Fig. 4d and Supplementary Fig. 14c). *Thbs4*^{KO/KO} mice showed significantly increased areas of magnetic susceptibility in the injured cortex as compared to the uninjured contralateral side, whereas control mice showed only well-defined hyperintense boundaries surrounding the resolving cortical lesions (Fig. 4d, f and Supplementary Fig. 14c, e). Although we did not detect these magnetic property differences early on at 2 d.p.i. between control and *Thbs4*^{KO/KO} mice (data not shown), the faster *R*₂^{*} rate and increased magnetic susceptibility in *Thbs4*^{KO/KO} mice at the later time point indicated presence of extensive, unresolved haemorrhage in their injured cortical regions.

Cortical vasculature compromise can be detected by using biotinylated fixable dextran as a tracer²⁶. We perfused mice 7 d.p.i. with biotinylated dextran and examined their injuries by fixed brain sections. IHC staining for GFAP⁺ astrocytes around the injury sites showed that whereas cortical glial scars can be readily identified in littermate

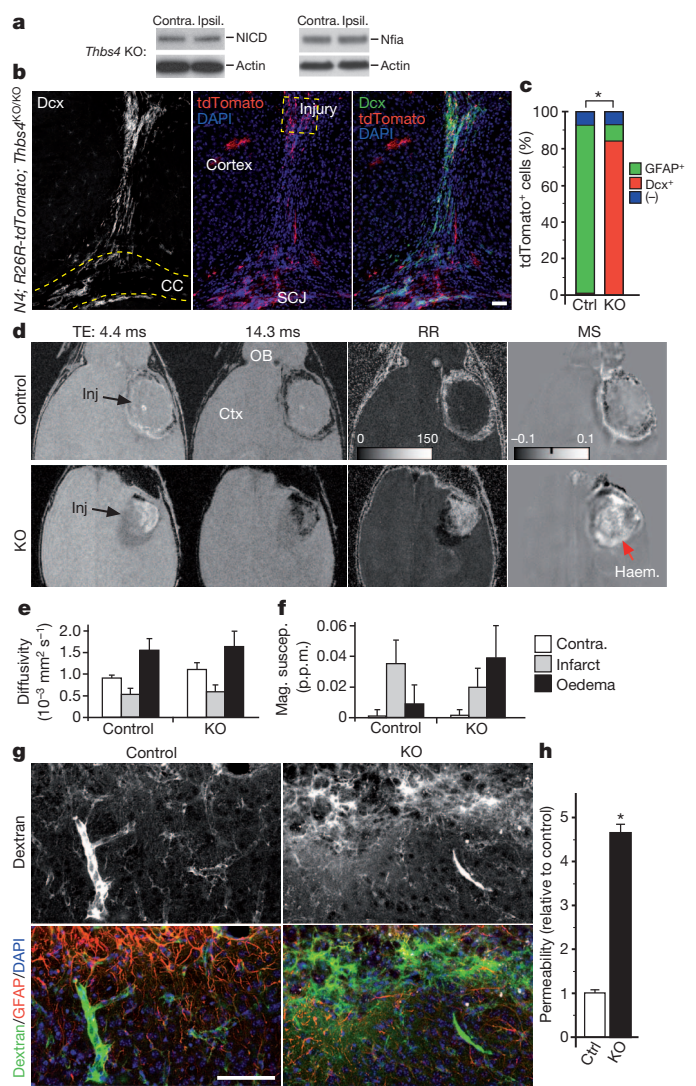


Figure 4 | SVZ astrogenesis defects in *Thbs4* mutant mice after cortical injury. **a**, Western blot analyses of NICD and Nfia protein levels in SVZ tissues collected 3 d.p.i. from *Thbs4* knockout mice, showing lack of upregulation in the ipsilateral versus contralateral SVZ. **b**, IHC staining for tdTomato showing Dcx expression from P6 tamoxifen-induced *N4; R26R-tdTomato; Thbs4*^{KO/KO} mice 14 d.p.i. Note robust Dcx⁺ cells at injury site. CC, corpus callosum (yellow dashed lines); SCJ, striatal–cortical junction. **c**, Quantification of total tdTomato⁺ cells around injury site 14 d.p.i., co-labelling with GFAP or Dcx. Lack of strong staining for either GFAP or Dcx was marked as (-). **P* < 0.002, *n* = 11 mice (control), 5 mice (KO), Wilcoxon rank sum test. **d**, MRI analyses of littermate pair, *Thbs4*^{KO/+} (control) and *Thbs4*^{KO/KO} (KO) 8 d.p.i. First and second panels show SPGR images, horizontal plane, at two echo times (TE, 4.4 and 14.3 ms). The third panel shows computed *R*₂^{*} relaxation rate (RR). The fourth panel shows corresponding magnetic susceptibility (MS). Hyperintense magnetic susceptibility indicates area of haemorrhage (Haem., red arrow). Scale bar units in RR = s⁻¹; magnetic susceptibility is in parts per million (p.p.m.). Ctx, cortex; Inj, injury site; OB, olfactory bulb. **e, f**, Quantitative measurements of mean diffusivity, magnetic susceptibility (Mag. suscep.), comparing contralateral cortex (Contra.) to areas of infarct and oedema caused by injury. Error bars indicate s.d. **g**, IHC staining of brain sections 7 d.p.i. to visualize GFAP⁺ astrocytes and biotinylated dextran infused through vasculature. Extravascular biotinylated dextran is readily seen around cortical injury site in knockout mice (close-ups from corresponding dashed boxes in Supplementary Fig. 12c). **h**, Quantification of parenchymal biotinylated dextran fluorescence next to injury site. **P* < 0.001, *n* = 5, Student's *t*-test; error bars indicate s.e.m. Scale bars: 50 μm (b, g).

controls, they were poorly formed/disorganized in *Thbs4*^{KO/KO} mutant mice (Fig. 4g and Supplementary Fig. 12c), consistent with their defects in post-injury SVZ astrogenesis. Furthermore, whereas biotinylated

dextran was largely retained inside the vasculature adjacent to injured areas in littermate controls, in *Thbs4*^{KO/KO} mice we detected widespread parenchymal biotinylated dextran next to the injury sites (Fig. 4g, h). These results are consistent with the MRI analyses above, and indicate continued microvascular bleeds after cortical injury.

Regulation of postnatal/adult astrogenesis from the SVZ niche is poorly understood. Here we have uncovered in mice that SVZ-derived astrocytes differ in *Thbs4* expression from their cortical counterparts. We found that cortical injury induces SVZ production of *Thbs4*^{hi} astrocytes, which migrate to the injury sites. Although it remains possible that *Thbs4* functions in multiple cell types after cortical injury, our findings are in agreement with protective roles for astrocytes during recovery process in the injured spinal cord²⁷. They revealed a surprising outcome where SVZ-mediated astrogenesis may be beneficial over neurogenesis in a period after cortical injury. Furthermore, we believe that these results will have important implications for therapeutic interventions using transplanted and/or endogenous NSCs after brain injury^{28,29}, as well as astrocytic tumours that can arise from the SVZ niche³⁰.

METHODS SUMMARY

Tissue preparation for IHC staining was as described¹¹. Images were acquired on Leica TCS SP5 confocal, with control and experimental samples imaged under identical settings. For injuries, rose bengal was delivered intraperitoneally (10 mg ml⁻¹ in saline, 0.1 mg g⁻¹ of body weight); light-induction: Zeiss KL1500 with 2.5 mm light-guide opening (light intensity at 5, 6-min duration). All mouse experiments were performed according to approved IACUC protocol at Duke University. For cultures, recombinant *Thbs4* (3 µg ml⁻¹) was freshly added to medium every other day. Protein extracts and western blotting were performed as described⁴, and NIH ImageJ software was used for quantifications. SVZ immunoprecipitation experiments were performed as described¹¹ with the following modification: CHAPS lysis buffer contained 5 mM CHAPS, 50 mM Tris (pH 7.4), 150 mM NaCl, 1 mM CaCl₂, 5% sucrose, 0.5% Triton X-100, protease inhibitor cocktail (Roche) plus sodium orthovanadate. To examine haemorrhage, 8-echo 3D spoiled-gradient-recalled (SPGR) sequence was used with parameters: field of view (FOV) = 22 × 11 × 11 mm³, matrix = 512 × 256 × 256, initial echo time (TE) = 4.4 ms, echo spacing = 4.9 ms, repetition time (TR) = 100 ms, flip angle = 45°. Mice were scanned on 9.4 Tesla (400 MHz) 89-mm vertical bore MRI scanner (Oxford Instruments) with shielded coil providing gradients of 2,200 mT m⁻¹. Biotinylated fixable dextran (10 kDa, 1.0 mg ml⁻¹, Invitrogen) experiments and their analyses were performed as described²⁶.

Full Methods and any associated references are available in the online version of the paper.

Received 10 January 2012; accepted 13 March 2013.

Published online 24 April 2013.

- Kriegstein, A. & Alvarez-Buylla, A. The glial nature of embryonic and adult neural stem cells. *Annu. Rev. Neurosci.* **32**, 149–184 (2009).
- Kelsch, W., Sim, S. & Lois, C. Watching synaptogenesis in the adult brain. *Annu. Rev. Neurosci.* **33**, 131–149 (2010).
- Ihrle, R. A. & Alvarez-Buylla, A. Lake-front property: a unique germinal niche by the lateral ventricles of the adult brain. *Neuron* **70**, 674–686 (2011).
- Paez-Gonzalez, P. *et al.* Ank3-dependent SVZ niche assembly is required for the continued production of new neurons. *Neuron* **71**, 61–75 (2011).
- Molofsky, A. V. *et al.* Astrocytes and disease: a neurodevelopmental perspective. *Genes Dev.* **26**, 891–907 (2012).
- Lawler, J. *et al.* Identification and characterization of thrombospondin-4, a new member of the thrombospondin gene family. *J. Cell Biol.* **120**, 1059–1067 (1993).
- Eroglu, C. The role of astrocyte-secreted matricellular proteins in central nervous system development and function. *J. Cell Commun. Signal.* **3**, 167–176 (2009).
- Deneen, B. *et al.* The transcription factor NFIA controls the onset of gliogenesis in the developing spinal cord. *Neuron* **52**, 953–968 (2006).
- Adams, J. C. & Lawler, J. The thrombospondins. *Cold Spring Harb. Perspect. Biol.* **3**, a009712 (2011).

- Liau, J. *et al.* Thrombospondins 1 and 2 are necessary for synaptic plasticity and functional recovery after stroke. *J. Cereb. Blood Flow Metab.* **28**, 1722–1732 (2008).
- Kuo, C. T. *et al.* Postnatal deletion of Numb/Numbl reveals repair and remodeling capacity in the subventricular neurogenic niche. *Cell* **127**, 1253–1264 (2006).
- Frolova, E. G. *et al.* Thrombospondin-4 regulates vascular inflammation and atherogenesis. *Circ. Res.* **107**, 1313–1325 (2010).
- Thored, P. *et al.* Persistent production of neurons from adult brain stem cells during recovery after stroke. *Stem Cells* **24**, 739–747 (2006).
- Givogri, M. I. *et al.* Notch signaling in astrocytes and neuroblasts of the adult subventricular zone in health and after cortical injury. *Dev. Neurosci.* **28**, 81–91 (2006).
- Li, L. *et al.* Focal cerebral ischemia induces a multilineage cytogenic response from adult subventricular zone that is predominantly gliogenic. *Glia* **58**, 1610–1619 (2010).
- Maxwell, K. A. & Dyck, R. H. Induction of reproducible focal ischemic lesions in neonatal mice by photothrombosis. *Dev. Neurosci.* **27**, 121–126 (2005).
- Aguirre, A., Rubio, M. E. & Gallo, V. Notch and EGFR pathway interaction regulates neural stem cell number and self-renewal. *Nature* **467**, 323–327 (2010).
- Morrison, S. J. *et al.* Transient Notch activation initiates an irreversible switch from neurogenesis to gliogenesis by neural crest stem cells. *Cell* **101**, 499–510 (2000).
- Han, H. *et al.* Inducible gene knockout of transcription factor recombination signal binding protein-J reveals its essential role in T versus B lineage decision. *Int. Immunol.* **14**, 637–645 (2002).
- Carlén, M. *et al.* Forebrain ependymal cells are Notch-dependent and generate neuroblasts and astrocytes after stroke. *Nature Neurosci.* **12**, 259–267 (2009).
- Meng, H., Zhang, X., Hankenson, K. D. & Wang, M. M. Thrombospondin 2 potentiates notch3/jagged1 signaling. *J. Biol. Chem.* **284**, 7866–7874 (2009).
- Yamamoto, S., Charnig, W. L. & Bellen, H. J. Endocytosis and intracellular trafficking of Notch and its ligands. *Curr. Top. Dev. Biol.* **92**, 165–200 (2010).
- Namihira, M. *et al.* Committed neuronal precursors confer astrocytic potential on residual neural precursor cells. *Dev. Cell* **16**, 245–255 (2009).
- Moseley, M. E., de Crespigny, A. J., Roberts, T. P., Kozniowska, E. & Kucharczyk, J. Early detection of regional cerebral ischemia using high-speed MRI. *Stroke* **24**, 160–165 (1993).
- Patel, M. R., Edelman, R. R. & Warach, S. Detection of hyperacute primary intraparenchymal hemorrhage by magnetic resonance imaging. *Stroke* **27**, 2321–2324 (1996).
- Daneman, R., Zhou, L., Kebede, A. A. & Barres, B. A. Pericytes are required for blood-brain barrier integrity during embryogenesis. *Nature* **468**, 562–566 (2010).
- Faulkner, J. R. *et al.* Reactive astrocytes protect tissue and preserve function after spinal cord injury. *J. Neurosci.* **24**, 2143–2155 (2004).
- Robel, S., Berninger, B. & Gotz, M. The stem cell potential of glia: lessons from reactive gliosis. *Nature Rev. Neurosci.* **12**, 88–104 (2011).
- Aboody, K., Capela, A., Niazi, N., Stern, J. H. & Temple, S. Translating stem cell studies to the clinic for CNS repair: current state of the art and the need for a Rosetta Stone. *Neuron* **70**, 597–613 (2011).
- Alcantara Llaguno, S. *et al.* Malignant astrocytomas originate from neural stem/progenitor cells in a somatic tumor suppressor mouse model. *Cancer Cell* **15**, 45–56 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Melton (Harvard) for *R26R-NICD* mice; T. Honjo (Kyoto) for *RBPjk-flox* mice; F. Wang for *R26R-tdTomato* mice; E. Rawlins (Cambridge) for *Foxj1-creER²* mice; B. Deneen (B.C.M.) and S. Singh for discussions; G. Lyons, R. Andersen, P. Heine, D. Fromme and S. Collins for project assistance; Duke Flow Cytometry Facility for help with FACS; W. Li and Duke Center for *In vivo* microscopy/brain imaging for MRI analyses; and T. Lechler, A. West and B. Hogan for comments on manuscript. This work was supported by National Biomedical Technology Resource Center Grant P41 RR005959 (C.L.) of P41 EB015897 to Duke Center for *In Vivo* Microscopy; George and Jean Brumley Endowment, Sontag Foundation, David and Lucile Packard Foundation, March of Dimes, and NIH Director's New Innovator Award 1 DP2 OD004453-01 (C.T.K.).

Author Contributions E.J.B. performed injury and biochemical experiments; D.L. performed gene expression and live-imaging experiments; K.A. performed *in vivo* immunoprecipitation experiments; P.P.-G. performed SVZ antibody staining and analyses; R.J., H.S. and D.S.W. assisted with injuries and their analyses; C.L. performed MRI scanning and quantitative analyses; C.E. provided reagents and experimental insight; C.T.K. performed transplantations and conceived the project. E.J.B., D.L. and R.J. assembled figures and C.T.K. wrote the paper. All authors discussed results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.T.K. (chay.kuo@duke.edu).

METHODS

Cell culture. Adherent SVZ primary cultures were isolated and grown as described³¹. Cortical astrocyte cultures were grown as described³². Jag1-Fc (R&D Systems) was plated at $5 \mu\text{g ml}^{-1}$ in PBS overnight to coat culture dish surfaces, followed by PBS washes. Recombinant Thbs4 ($3 \mu\text{g ml}^{-1}$, C. Eroglu), DAPT ($5 \mu\text{M}$, Tocris) and DBZ (5 nM , Millipore) were freshly added to culture medium every other day during *in vitro* differentiation where indicated. Dynasore³³ (Sigma) was used at $30 \mu\text{M}$ in culture medium.

Immunohistochemistry and live-imaging analyses. Preparation of brain tissue for immunohistochemistry (IHC) was as described¹¹. For Thbs4 and DAB staining: after transcardiac perfusion, brains were removed, postfixed overnight and cryoprotected in 30% sucrose at 4°C . $30 \mu\text{m}$ coronal sections were serially cut, and immediately incubated for 1 h at room temperature, floating in PBST blocking buffer containing 10% donkey serum. Primary antibody incubation was carried out at 4°C in blocking buffer overnight, followed by washes in PBST $\times 3$, PBS alone $\times 3$, and secondary antibody incubation in blocking buffer for 2 h at room temperature. DAB staining was carried out according to manufacturer instructions (Vectastain ABC kit, Vector Labs). For cell counting and quantifications, IHC-stained coronal sections, $90 \mu\text{m}$ apart, starting at the anterior portion of the anterior commissure through the septofimbrial nuclei, were analysed and counted. Primary antibodies against the following antigens were used: Thbs4 (goat, 1:200, R&D Systems), GFP (chicken, 1:500, Aves Labs), RFP (rabbit, 1:1,000, Rockland), GFAP (mouse, 1:2,000, Sigma), Mash1 (mouse, 1:100, BD Biosci), Dcx (guinea-pig, 1:1,000, Millipore), NeuN (mouse, 1:500, Millipore), Notch1 (rabbit monoclonal EP1238Y, 1:1,000, Epitomics), Olig2 (rabbit, 1:800, Millipore), NG2 (mouse, 1:200, R&D Systems). Biotinylated fixable dextran (molecular mass of 10 kDa , 1.0 mg ml^{-1} , Invitrogen) experiments and their image analyses were performed as described²⁶. Live imaging of migrating neuroblasts was performed as described³⁴. All fluorescent images were acquired on Leica TCS SP5 confocal microscope, with control and experimental samples imaged under identical settings.

FACS sorting and gene expression analyses. Cortical and SVZ tissues from GFAP-GFP mice (JAX 003257) were dissected in warm NSC culture media³¹, and dissociated with neural tissue dissociation kit (Miltenyi) according to the manufacturer's protocol. Cell suspension was added to FACS tube on ice, sorted on BD FACS DiVa sorter, followed immediately by Trizol RNA extraction. 500 ng of total RNA was used for cDNA synthesis using SuperScript VILO cDNA synthesis kit (Invitrogen). Quantitative PCR analyses were performed as described³⁵, using the following DNA primers: *Thbs4* forward $5'$ -ATCCCTGCTATCCAGG TGTG- $3'$, reverse $5'$ -GGCAGCTCCTTCAGTCTTG- $3'$; *Nfia* forward $5'$ -CCA GCCAGCAAGTGAAG- $3'$, reverse $5'$ -GCTCAGTCACACTGAAACACC- $3'$; housekeeping gene controls *Gapdh* forward $5'$ -CATGGCCTCCGTGTTC CT- $3'$, reverse $5'$ -TGATGTCATCATACTGGCAGGTT- $3'$; and *Ppia* forward $5'$ -CGAGCTCTGACACTGGAG- $3'$, reverse $5'$ -GATGCCAGGACCTGTATG CT- $3'$. Consistent results were obtained using either housekeeping primer sets.

SDS-PAGE and immunoblotting. Protein extract preparation and western blotting were performed as described⁴. NIH ImageJ software was used for quantification analyses. Primary antibodies against the following antigens were used: Thbs4 (goat, 1:1,000, R&D Systems), Thbs2 (mouse, 1:500, Millipore), Dcx (rabbit, 1:1,000, Cell Signaling), NICD/Notch1 (rabbit EP1238Y, 1:1,000, Millipore), Nfia (rabbit, 1:2,000, Active Motif), and actin (mouse, 1:5,000, Abcam). For pull-down experiments, purified recombinant Thbs4 or Thbs2 (R&D Systems) was incubated for 1 h at 37°C with either 200 ng mouse Notch1-Fc chimaera (R&D Systems, 5267-TK) or mouse IgG2a isotype control (Abcam) in Dulbecco's PBS + 1% BSA, 0.1% Triton X-100. Protein G-coupled DynaBeads (Invitrogen), pre-blocked in DPBS + 3% BSA, were then added for 10 min at 37°C , followed by five washes in DPBS + 0.1% Triton X-100, and analysed by western blotting. For cultured cells, media with or without Thbs4 ($3 \mu\text{g ml}^{-1}$) were first added for 4 h, washed two times with cold PBS, followed by addition of lysis buffer (50 mM Tris pH 8.0, 75 mM NaCl, 1 mM EDTA, 5% sucrose, 0.25% Triton X-100, mini-complete protease inhibitor cocktail), sonication, and centrifuged for 10 min at 4°C . Lysates were pre-cleared with Protein A agarose beads (Roche). Rabbit anti-Notch1 antibody ($5 \mu\text{g}$, Millipore EP1238Y) was added to 1 ml of lysate and incubated overnight at 4°C , followed by addition of Protein A agarose beads, washes as described above, and western blotting analysis. SVZ

immunoprecipitation experiments were performed as described¹¹ with following modifications: CHAPS lysis buffer contained 5 mM CHAPS, 50 mM Tris (pH 7.4), 150 mM NaCl, 1 mM CaCl₂, 5% sucrose, 0.5% Triton X-100, protease inhibitor cocktail (Roche) plus sodium orthovanadate. Supernatants after lysis were subject to two rounds of pre-clearing with BSA-blocked Protein G agarose beads (Roche) for 1 h each, followed by incubation with rabbit anti-Notch1 antibody (Millipore EP1238Y) for 1 h, and overnight incubation with BSA-blocked Protein G agarose beads, all at 4°C . Beads were washed five times in lysis buffer, three times in PBS before suspension in sample buffer for western blotting.

Cortical injury and *in vivo* injections. Photothrombosis cortical injury was performed as described³⁶. Briefly, mice were anaesthetized and body temperature maintained at 37°C with recirculating water heating pad. Saline solution of rose bengal photosensitive dye (10 mg ml^{-1} in saline, 0.1 mg g^{-1} of body weight) was delivered intraperitoneally. Midline scalp incision was made to expose the skull. As an external light source, Zeiss KL 1500 LCD (light intensity at 5, 6-min duration) with 2.5-mm opening light guide was used to induce photothrombosis. Transplantation of passage 2 primary adherent SVZ culture was performed as described⁴. 13 out of 37 transplanted animals showed successfully grafted tdTomato⁺ cells on analyses. Tamoxifen (10 mg ml^{-1} , freshly dissolved in corn oil) was injected intraperitoneally at 0.15 mg g^{-1} of body weight to induce CreER-mediated recombination. All mouse experiments were performed according to approved protocol by IACUC at Duke University.

Magnetic resonance imaging and analyses. Pairs of injured *Thbs4*^{KO/+} and *Thbs4*^{KO/KO} mice were scanned on 9.4 Tesla (400 MHz) 89-mm vertical bore MRI scanner (Oxford Instruments) with shielded coil providing gradients of $2,200 \text{ mT m}^{-1}$. Mice were perfused with 10% buffered formalin as previously described³⁷, followed by overnight fixation in formalin and 3-day incubation in PBS before imaging. All brains were kept within the cranium to prevent potential damage by removal. Diffusion-weighted images (DWI) were acquired using three-dimensional spin-echo sequence with following parameters: field of view (FOV) = $22 \times 11 \times 11 \text{ mm}^3$, matrix = $164 \times 82 \times 82$, TE = 12 ms, TR = 2,500 ms. One image volume was acquired without diffusion weighting, and diffusion weighting was achieved by applying two half-sine gradient pulses around the 180° refocusing pulse. To produce isotropic weighting, six non-collinear diffusion encoding directions were used at b -value = $3,000 \text{ s mm}^{-2}$. A diffusion tensor was fitted using seven image volumes and mean diffusivity was calculated as described³⁸. To examine intracranial haemorrhage, brains were scanned using 8-echo three-dimensional spoiled-gradient-recalled (SPGR) sequence with the following parameters: FOV = $22 \times 11 \times 11 \text{ mm}^3$, matrix = $512 \times 256 \times 256$, initial TE = 4.4 ms, echo spacing = 4.9 ms, TR = 100 ms, flip angle = 45° . R_2^* relaxation rate was fitted with exponential decay curve. Tissue magnetic susceptibility was quantified as described³⁹. Regions of infarct and oedema were segmented out based on mean diffusivity maps generated via FSL software (Oxford University), and comparable regions in the contralateral uninjured cortex were drawn as references. For each region, mean diffusivity, R_2^* relaxation rate and magnetic susceptibility were measured. Similar results were obtained with addition of MRI contrast agent ProHance (Bracco Diagnostics) to formalin fixation (1:9, v:v).

- Scheffler, B. *et al.* Phenotypic and functional characterization of adult brain neurogenesis. *Proc. Natl Acad. Sci. USA* **102**, 9353–9358 (2005).
- Christopherson, K. S. *et al.* Thrombospondins are astrocyte-secreted proteins that promote CNS synaptogenesis. *Cell* **120**, 421–433 (2005).
- Macia, E. *et al.* Dynasore, a cell-permeable inhibitor of dynamin. *Dev. Cell* **10**, 839–850 (2006).
- Platel, J. C. *et al.* NMDA receptors activated by subventricular zone astrocytic glutamate are critical for neuroblast survival prior to entering a synaptic network. *Neuron* **65**, 859–872 (2010).
- McDowell, K. A. *et al.* Reduced cortical BDNF expression and aberrant memory in Carf knock-out mice. *J. Neurosci.* **30**, 7453–7465 (2010).
- Lee, J. K. *et al.* Photochemically induced cerebral ischemia in a mouse model. *Surg. Neurol.* **67**, 620–625 (2007).
- Johnson, G. A., Cofer, G. P., Gewalt, S. L. & Hedlund, L. W. Morphologic phenotyping with MR microscopy: the visible mouse. *Radiology* **222**, 789–793 (2002).
- Basser, P. J., Mattiello, J. & LeBihan, D. MR diffusion tensor spectroscopy and imaging. *Biophys. J.* **66**, 259–267 (1994).
- Li, W., Wu, B. & Liu, C. Quantitative susceptibility mapping of human brain reflects spatial variation in tissue composition. *Neuroimage* **55**, 1645–1656 (2011).

X-ray phase-contrast *in vivo* microtomography probes new aspects of *Xenopus* gastrulation

Julian Moosmann¹, Alexey Ershov^{1,2}, Venera Altapova³, Tilo Baumbach^{1,3}, Maneeshi S. Prasad⁴, Carole LaBonne⁴, Xianghui Xiao⁵, Jubin Kashef⁶ & Ralf Hofmann¹

An ambitious goal in biology is to understand the behaviour of cells during development by imaging—*in vivo* and with subcellular resolution—changes of the embryonic structure. Important morphogenetic movements occur throughout embryogenesis, but in particular during gastrulation when a series of dramatic, coordinated cell movements drives the reorganization of a simple ball or sheet of cells into a complex multi-layered organism¹. In *Xenopus laevis*, the South African clawed frog and also in zebrafish, cell and tissue movements have been studied in explants^{2,3}, in fixed embryos⁴, *in vivo* using fluorescence microscopy^{5,6} or microscopic magnetic resonance imaging⁷. None of these methods allows cell behaviours to be observed with micrometre-scale resolution throughout the optically opaque, living embryo over developmental time. Here we use non-invasive *in vivo*, time-lapse X-ray microtomography, based on single-distance phase contrast and combined with motion analysis, to examine the course of embryonic development. We demonstrate that this powerful four-dimensional imaging technique provides high-resolution views of gastrulation processes in wild-type *X. laevis* embryos, including vegetal endoderm rotation, archenteron formation, changes in the volumes of cavities within the porous interstitial tissue between archenteron and blastocoel, migration/confrontation of mesendoderm and closure of the blastopore. Differential flow analysis separates collective from relative cell motion to assign propulsion mechanisms. Moreover, digitally determined volume balances confirm that early archenteron inflation occurs through the uptake of external water. A transient ectodermal ridge, formed in association with the confrontation of ventral and head mesendoderm on the blastocoel roof, is identified. When combined with perturbation experiments to investigate molecular and biomechanical underpinnings of morphogenesis, our technique should help to advance our understanding of the fundamentals of development.

Gastrulation is a morphogenetic milestone in which complex cell and tissue rearrangements position the embryonic germ layers. It has been studied in living and fixed gastrulae using time-lapse light microscopy^{8–11}, and in fixed, cleaved or dissected embryos³ by electron microscopy. However, these two-dimensional (2D) imaging methods are limited in their capacity to reveal deep tissue dynamics. As a consequence, certain gastrulation movements, such as convergent extension of the mesoderm² and vegetal rotation of the endoderm¹², have been analysed *in vitro*. Although studies of explants provide valuable insights, they are unlikely to fully recapitulate *in vivo* morphogenetic behaviours which depend on cell and tissue interactions within the intact embryo^{1,13,14}. Confocal light microscopy¹⁵ and digital scanned light-sheet microscopy^{6,16} have been used to study *in vivo* developmental processes in four dimensions (4D). However, both techniques are most applicable to optically translucent embryos (for example, zebrafish) and require fluorescence labelling of selected subcellular structures^{15,17}. *In vivo* microscopic magnetic

resonance imaging has been used to investigate opaque embryos⁷ (for example, mice and *X. laevis*), but this technique is unable to resolve structure at the cellular level.

Here we use X-ray phase-contrast microtomography to overcome these limitations and to avoid high dose depositions introduced by conventional X-ray absorption imaging¹⁸. Exploiting the interference of electromagnetic waves, this imaging technique benefits from their high coherence and intensity when produced by modern synchrotrons. Early developmental stages of vertebrate embryos are composed of light chemical elements. As a consequence, phase shifts in the transmitted wave field dominate attenuation by three orders of magnitude for hard X-rays^{19–21}. This facilitates *in vivo* X-ray imaging. Due to a high penetration depth, phase-contrast microtomography represents a genuine 3D imaging modality, and it visualizes both cell shape and subcellular structures such as nuclei and yolk platelets (Supplementary Video 1)^{22–24}.

Radiation dose induces heat load and radiolysis of water in living samples. Accordingly, our set-up (Fig. 1 and Methods) was configured to minimise dose deposition while maintaining sufficient image contrast at an optimal X-ray energy $E \sim 30$ keV. Under these conditions heat load

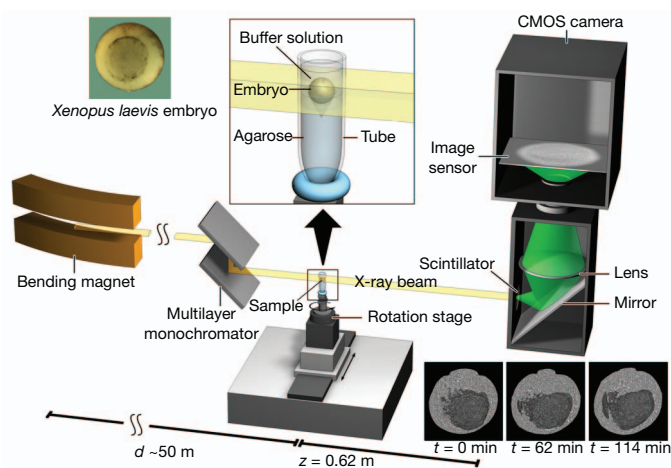


Figure 1 | Experimental set-up for propagation-based phase-contrast X-ray microtomography. A quasi-parallel photon beam is generated from a synchrotron electron beam traversing the field of a bending magnet. After beam shaping and monochromatization, X-ray wave fronts propagate over a distance d (~ 50 m) to impinge on the sample (living *X. laevis* embryo immersed in buffer solution and suspended by agarose) mounted on a rotation stage for tomographic data acquisition. The 2D detector, at a distance $z = 62$ cm behind the sample, consists of a scintillator, converting X-rays into visible light, followed by a mirror, a lens, and a complementary metal oxide semiconductor (CMOS) camera (effective pixel size $\Delta x = 2.2 \mu\text{m}$).

¹Institute for Photon Science and Synchrotron Radiation, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, D-76344 Eggenstein-Leopoldshafen, Germany. ²Institute of Physics and Technology, Department of General Physics, National Research Tomsk Polytechnic University, Lenin Avenue 30, 634050 Tomsk, Russia. ³Laboratory for Applications of Synchrotron Radiation, Karlsruhe Institute of Technology, Postfach 6980, D-76128 Karlsruhe, Germany. ⁴Department of Molecular Biosciences, Northwestern University, 2205 Tech Drive, Hogan 2-100, Evanston, Illinois 60208-1230, USA. ⁵Advanced Photon Source, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA. ⁶Zoological Institute II, Karlsruhe Institute of Technology, Cell and Developmental Biology, Kaiserstrasse 12, D-76131 Karlsruhe, Germany.

is predicted to be negligible (Supplementary Fig. 1 and Supplementary Information), which was confirmed by experiment (Supplementary Fig. 2). Radiolysis of water eventually leads to accumulation of lethal radicals resulting in sudden dissolution of tissue structure (Supplementary Video 16). However, before this point, morphogenetic processes and embryonic viability appear unimpaired. For a direct assay of viability, an easily disrupted process—blastopore closure—was examined. Closure rates were found to be comparable in X-rayed embryos (Supplementary Videos 3, 4 and 6) and control gastrulae viewed by light microscopy (Supplementary Fig. 3 and Supplementary Video 13).

Additional time-lapse *in vivo* scans, including neurulation, were acquired (Supplementary Videos 5–10 and Supplementary Information). These sequences confirm normal development in X-rayed embryos, even with harsher irradiation conditions than those used to acquire the gastrulation sequence analysed below (Supplementary Video 3).

Figure 2a–c shows mid-sagittally halved tomographic renderings of an embryo at developmental stages 11.5, 12.0 and 12.5 (Supplementary Video 3). Specific embryonic structures are discernible, including the blastopore, Brachet's cleft, archenteron, ventral and dorsal blastopore lips, blastocoel, blastocoel roof and floor, and the porous, interstitial tissue between archenteron and blastocoel ('pipe' system, Supplementary Video 2). Noteworthy processes captured include the progression of epiboly by a thinning of ectoderm, blastopore closure and archenteron formation (Supplementary Videos 3 and 12).

To analyse cell movements, we computed 3D velocity fields $\vec{v} = (v_x, v_y, v_z)$ from the time-lapse series of reconstructed volumes employing optical flow methods²⁵ adapted to tomographic data. Essentially, the latter determine the displacement of a certain voxel in a given volume towards the associated voxel in the successive volume such that grey-value differences are minimised under certain constraints (Supplementary Information). Morphogenesis is the collective consequence of hierarchical, diverse, and coupled tissue/cell movements, all of which

are captured by field \vec{v} . Therefore, we further explored (1) the collective aspects of \vec{v} , (2) its spatially differential behaviour and (3) its implications for individual cell trajectories. Velocity (flow) fields, depicted in Fig. 2g–i, reveal collective cell movements within the gastrula for several morphogenetic processes, including rotation of the vegetal endoderm (white arrowhead in Fig. 2g, Supplementary Video 11)¹², involution of the mesendoderm at the dorsal and ventral blastopore lip (white arrows in Fig. 2g), and migration of the mesendoderm on the blastocoel roof towards the ventral animal pole (red arrowhead in Fig. 2g)¹. At stage 11.5, involuting cells are fastest ($\sim 4 \mu\text{m min}^{-1}$) at the dorsal and ventral inner blastopore lips (Fig. 2g). By stage 12, prominent horizontal motion of the endoderm is observed on the ventral side (Fig. 2h and Supplementary Video 11). In addition, rapid migration of presumptive leading-edge head mesendodermal cells on the blastocoel roof, a bulk movement of the mesendoderm towards the animal pole, and localized cell displacements (archenteron formation) are discernible. Once the dorsal and ventral edges of the mesendodermal mantle make contact and enclose the blastocoel, an overall slowing of cell movements is observed on the ventral side (Fig. 2i, stage 12.5). Also, a strong inward displacement of cells in the anterior archenteron wall occurs, probably reflecting the rapid expansion of this cavity taking place towards the end of gastrulation²⁶. Next, differential cell and tissue movements were examined through field

$G \equiv |\vec{\nabla} v_x| + |\vec{\nabla} v_y| + |\vec{\nabla} v_z| \quad \vec{\nabla} \equiv \vec{e}_x \frac{\partial}{\partial x} + \vec{e}_y \frac{\partial}{\partial y} + \vec{e}_z \frac{\partial}{\partial z}$ (the spatial gradient operator; $\vec{e}_x, \vec{e}_y, \vec{e}_z$ an orthonormal basis). Small values of G indicate collective motion, large values capture gradients in the velocity magnitude $|\vec{v}|$ and/or variations in the direction of \vec{v} . Unlike the field $S \equiv |\vec{\nabla} \cdot \vec{v}|$, G is rotationally non-invariant and thus depends on the choice of Cartesian coordinates. However, G in contrast to S captures directional deviations at constant magnitude $|\vec{v}|$, and therefore we favoured it here. Figures 3a, b and 3c, d depict $|\vec{v}|$ and G , respectively. In computing \vec{v} , optical flow is optimized to feature large, high-contrast mesendodermal or endodermal cells, thereby marginalizing the visualization of posterior involution movements and radial cell intercalation in the ectoderm. Contour C encloses a large, collectively moving cell mass (Fig. 3a), extending from the vegetal region (C1, Fig. 3c) along the posterior–anterior axial mesendoderm and endoderm on the dorsal side (C2, Fig. 3c). The movement of cells within C1 is essentially collective (small values of G), driven mainly by blastopore closure, whereas within C2, a collective-motion component of similar velocity magnitude is due to involution of the dorsal mesendoderm (transition from C1 to C2 in Fig. 3c). In addition, relative cell motion occurs within C2 that is attributable to mediolateral intercalation associated with convergent extension. This explains the acceleration (Fig. 2g, h) of the anterior cells within C (Fig. 3a) towards the animal pole, thus filling the gap (C3, Fig. 3a) in velocity magnitude (compare Fig. 3a and 3b). This gap indicates active migration²⁷ of the leading-edge cells in the small region A (Fig. 3a, b) because their motion cannot be powered by the cell mass within C2, and because A cannot convey sizable convergent extension effects. Individual cells (I1) crawl on the blastocoel floor (Fig. 3c, d, see also Supplementary Video 4), and a single cell (I2) migrates on the endodermal bulk (Fig. 3d). Finally, trajectories of adjacent cells are computed by time integration of 3D flow fields in order to analyse differences in their dynamics (Fig. 3e, Supplementary Information and Supplementary Fig. 4).

We proceeded with an analysis of morphological changes. Archenteron formation is among the least understood aspects of gastrulation. In the mid-horizontal slice of Fig. 2e, two distinct, hollow regions occur (black arrows) which appear to merge (Fig. 2f). However, rendering of the entire volume (Supplementary Videos 12 and 14) reveals a heart-shaped and contiguous archenteron (Fig. 3f, g), highlighting the importance of 3D imaging. The increased porosity of the 'pipe' system may suggest (Fig. 3f, g) that archenteron inflation is driven by fluid transfer from the blastocoel⁴. To assess this we segmented blastocoel, archenteron, cavities in the 'pipe' system, and the entire embryo at different

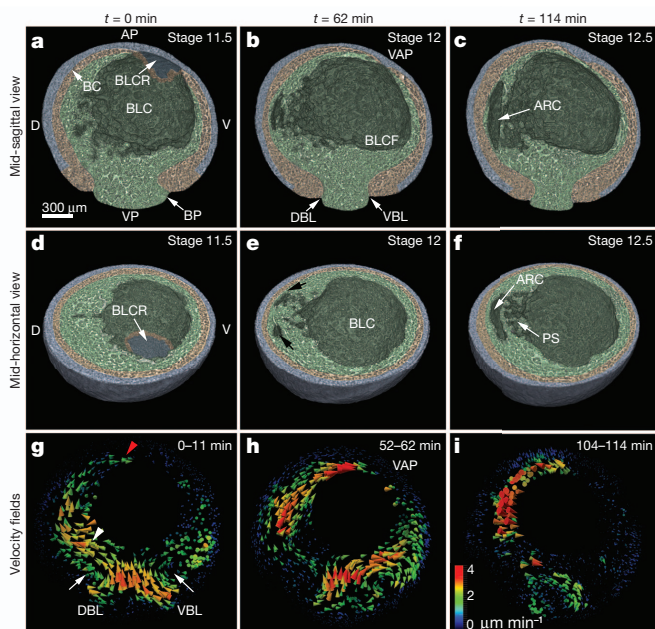


Figure 2 | 3D time-lapse series of *X. laevis* embryo during mid-gastrulation. a–f, Mid-sagittally (a–c) and mid-horizontally (d–f) halved embryo renderings at stages 11.5 (0 min), 12 (62 min) and 12.5 (114 min). Ectoderm (blue), mesoderm (orange), and endoderm (green). g–i, Velocity fields on a 180- μm -thick 3D slab centred about the cutting planes of (a–c). Colour bar indicates velocity magnitude representation. Animal pole (AP), Archenteron (ARC), Brachet's cleft (BC), blastocoel (BLC), blastocoel floor (BLCF), blastocoel roof (BLCR), blastopore (BP), dorsal and ventral sides (D, V), dorsal and ventral blastopore lip (DBL, VBL), 'pipe' system in-between archenteron and blastocoel (PS), ventral animal pole (VAP) and vegetal pole (VP).

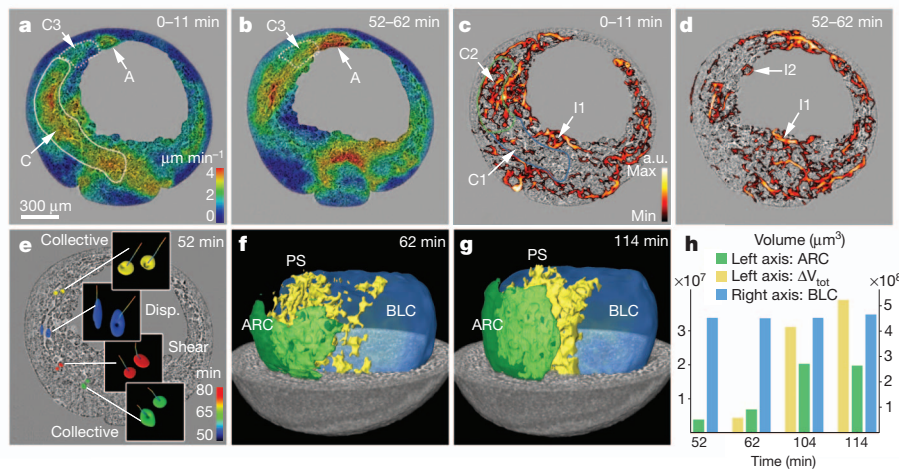


Figure 3 | Collective versus differential flow and cavity morphogenesis. **a, b**, Magnitude $|\vec{v}|$ of velocity on sagittal slice for two different times. **c, d**, Field G on same sagittal slice and for the same times. **e**, Sagittal slice at 52 min, highlighted cell pairs line the archenteron, associated trajectories (period of

30 min) (colour bar: blue, early; red, late). **f, g**, 3D renderings of cavities within the 'pipe' system in between archenteron and blastocoel for times 62 min and 114 min. **h**, Volume changes of ARC, BLC and gastrula from 52 min to 114 min.

times (Fig. 3f–h and Supplementary Video 14). Strikingly, although the archenteron inflated, the blastocoel volume remained constant (Fig. 3h), whereas the volumes of the 'pipe' system and the entire embryo increased. This indicates that early archenteron expansion is not driven by fluid transfer from the blastocoel, but instead by the uptake of external water²⁸. Indeed, this is implied by the conservation of cell–mass volume during gastrulation²⁸, our results concerning volume changes (archenteron, 'pipe' system and entire embryo) (Fig. 3f–h), and the absence of additional cavities in the embryonic fluid balance (for example, disappearance of small blastopore cavities well before archenteron inflation, Supplementary Video 17 and Supplementary Information).

An interesting aspect of gastrulation is the active, non-invasive migration of mesendodermal cells on the ectoderm of the blastocoel roof. A model for this was proposed recently in which cycles of Eph/ephrin-dependent attachment and repulsion drive cell migration²⁷. Our data capture, in detail, migration of the leading-edge head and ventral mesendoderm on their ectodermal substrate (Fig. 4f). The sagittal

slices in Fig. 4a–e depict regions of contact and non-contact between the mesendoderm and the ectoderm, creating cavities consistent with that model²⁷. Moreover, the ectoderm becomes transiently deformed towards the centre of the blastocoel, forming a cusp where the dorsal and ventral mesendoderm meet and make contact (Fig. 4b, c and Supplementary Video 15). The epiboly or spread of ectodermal cells over the surface of the gastrula runs counter to this. In 3D, the cusp of Fig. 4b, c corresponds to a ridge (Fig. 4g) that may represent a local contraction of the ectoderm itself, or arises from a local, adhesive pull by the dorsal and ventral mesendoderm. When these two eventually overlap, their contact with the ectoderm is lost, relaxing the cusp (Fig. 4c–e and Supplementary Video 18). In turn, new adhesive interactions between mesendodermal cells were observed (Fig. 4c). The formation of this ridge in explants is unlikely, owing to disruption of tissue tension; to our knowledge, it has not been seen before.

The present study demonstrates that propagation-based, single-distance phase-contrast X-ray microtomography, combined with optical flow analysis, is a useful tool for 4D *in vivo* investigation of *X. laevis* embryonic development. Analysis of differential motion permits active or passive tissue and cell movements to be distinguished, providing insights into how collective motion (convergent extension, collective cell migration on bulk tissue) is powered by individual cell behaviour. Time-lapse series with subcellular resolution can be captured over periods of two hours without adverse effects on development and with an option for longer observation times at lower resolution. When used in high-throughput perturbation experiments, we expect this imaging modality to advance our understanding of the molecular mechanisms and biomechanical processes driving morphogenesis.

METHODS SUMMARY

The time-lapse series analysed in this work relies on phase-contrast X-ray microtomography performed at beamline station 2-BM-B of Advanced Photon Source (monochromatization of $\Delta E/E = 10^{-2}$, flux density of $\sim 10^{12}$ photons per second per mm^2 , hutch temperature $\sim 25^\circ\text{C}$). The experimental set-up was optimized for low dose deposition (15 ms exposure per projection, $N = 1,200$ projections, 18 s to record one tomogram with continuous rotation of sample) at sufficient image contrast, avoiding blurring effects due to cell motion and source size ($3.4\ \mu\text{m}$ horizontal image blurring at $z = 62\ \text{cm}$, horizontal coherence length of $\sim 4.2\ \mu\text{m}$). Subsequent to phase retrieval^{22,24} a filtered back-projection algorithm²⁹ was applied to compute the 3D distribution of electron density (Supplementary Video 1 and Supplementary Information). The waiting time between tomographic recordings was 10 min, sufficiently short to perform 3D optical flow analysis (Supplementary Information). *In vitro* fertilisation, embryo culture and staging were carried out as described³⁰. All experiments were performed in accordance with a protocol approved by Northwestern University's Committee on Animal Care and Use (Animal assurance number A328301).

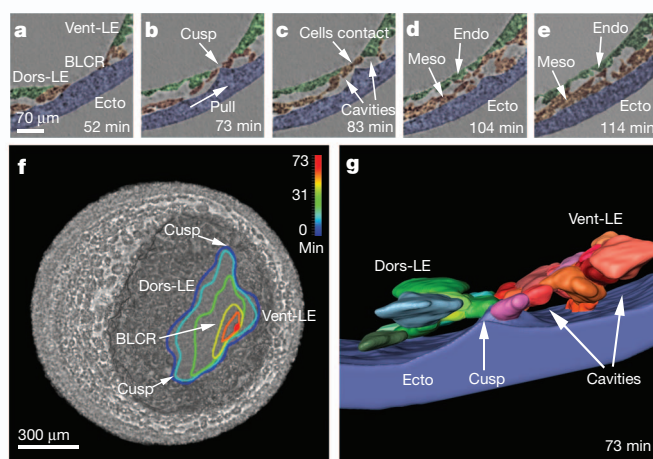


Figure 4 | Confrontation of head and ventral mesendoderm. **a–e**, Sequence of sagittal slices through confrontation zone at 52, 73, 83, 104 and 114 min showing formation and relaxation of a cusp of ectoderm on the blastocoel roof. Dorsal and ventral leading edges (Dors-LE, Vent-LE). Ectoderm (blue), mesoderm (orange), endoderm (green). **f**, Sequence of shrinking contours (projections of migrating leading edge along posterior–anterior axis onto horizontal plane) starting at 0 min (blue), terminating at 73 min. **g**, 3D rendering of individual cells and cavities in the confrontation zone at 73 min. Dors-LE cells (shades of green), Vent-LE cells (shades of red).

Full Methods and any associated references are available in the online version of the paper.

Received 7 November 2012; accepted 21 March 2013.

- Keller, R., Davidson, L. A. & Shook, D. R. How are we shaped: The biomechanics of gastrulation. *Differentiation* **71**, 171–205 (2003).
- Keller, R. & Danilchik, M. Regional expression, pattern and timing of convergence and extension during gastrulation of *Xenopus laevis*. *Development* **103**, 193–209 (1988).
- Damm, E. W. & Winklbauer, R. PDGF-A controls mesoderm cell orientation and radial intercalation during *Xenopus* gastrulation. *Development* **138**, 565–575 (2011).
- Ewald, A. J., Peyrot, S. M., Tyska, J. M., Fraser, S. E. & Wallingford, J. B. Regional requirements for dishevelled signaling during *Xenopus* gastrulation: separable effects on blastopore closure, mesendoderm internalization, and archenteron formation. *Development* **131**, 6195–6209 (2004).
- Huisken, J. & Stanier, D. Y. R. Selective plane illumination microscopy techniques in developmental biology. *Development* **136**, 1963–1975 (2009).
- Keller, Ph. J., Schmidt, A. D., Wittbrodt, J. & Stelzer, E. H. K. Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. *Science* **322**, 1065–1069 (2008).
- Papan, C., Velan, S. S., Fraser, S. E. & Jacobs, R. E. 3D time-lapse analysis of *Xenopus* gastrulation movements using μ MRI. *Dev. Biol.* **235**, 189 (2001).
- Wallingford, J. B., Fraser, S. E. & Harland, R. M. Convergent extension: The molecular control of polarized cell movement during embryonic development. *Dev. Cell* **2**, 695–706 (2002).
- Hardin, J. & Keller, R. The behaviour and function of bottle cells during gastrulation of *Xenopus laevis*. *Development* **103**, 211–230 (1988).
- Weisz, P. B. The normal stages in the development of the South African clawed toad, *Xenopus laevis*. *Anat. Rec.* **93**, 161–169 (1945).
- Ewald, A. J., McBride, H., Reddington, M., Fraser, S. E. & Kerschmann, R. Surface imaging microscopy, an automated method for visualizing whole embryo samples in three dimensions at high resolution. *Dev. Dyn.* **225**, 369–375 (2002).
- Winklbauer, R. & Schürfeld, M. Vegetal rotation, a new gastrulation movement involved in the internalization of the mesoderm and endoderm in *Xenopus*. *Development* **126**, 3703–3713 (1999).
- Keller, R., Shook, D. & Skoglund, P. The forces that shape embryos: physical aspects of convergent extension by cell intercalation. *Phys. Biol.* **5**, 015007 (2008).
- Blanchard, G. B. & Adams, R. Measuring the multi-scale integration of mechanical forces during morphogenesis. *Curr. Opin. Genet. Dev.* **21**, 653–663 (2011).
- Ruffins, S. W., Russells, E. J. & Fraser, S. E. Towards a Tralfamadorian view of the embryo: multidimensional imaging of development. *Curr. Opin. Neurobiol.* **12**, 580–586 (2002).
- Greger, K., Swoger, J. & Stelzer, E. H. K. Basic building units and properties of a fluorescence single plane illumination microscope. *Rev. Sci. Instrum.* **78**, 023705 (2007).
- Dubertret, B. *et al.* *In vivo* imaging of quantum dots encapsulated in phospholipid micelles. *Science* **298**, 1759–1762 (2002).
- Cheng, K. C., Xin, X., Clark, D. P. & La Riviere, P. Whole-animal imaging, gene function and the Zebrafish phenome project. *Curr. Opin. Genet. Dev.* **21**, 620–629 (2011).
- Snigirev, A., Snigireva, I., Kohn, V., Kuznetsov, S. & Schelokov, I. On the possibilities of X-ray phase contrast microimaging by coherent high-energy synchrotron radiation. *Rev. Sci. Instrum.* **66**, 5486–5492 (1995).
- Wilkins, S. W., Gureyev, T. E., Gao, D., Pogany, A. & Stevenson, A. W. Phase-contrast imaging using polychromatic hard X-rays. *Nature* **384**, 335–338 (1996).
- Nugent, K. A., Gureyev, T. E., Cookson, D. F., Paganin, D. M. & Barnea, Z. Quantitative phase imaging using hard X rays. *Phys. Rev. Lett.* **77**, 2961–2964 (1996).
- Paganin, D. & Nugent, K. A. Noninterferometric phase imaging with partially coherent light. *Phys. Rev. Lett.* **80**, 2586–2589 (1998).
- Cloetens, P. *et al.* Hard X-ray phase imaging using simple propagation of a coherent synchrotron radiation beam. *J. Phys. D Appl. Phys.* **32**, A145–A151 (1999).
- Hofmann, R., Moosmann, J. & Baumbach, T. Criticality in single-distance phase retrieval. *Opt. Express* **19**, 25881–25890 (2011).
- Brox, T., Bruhn, A., Papenberg, N. & Weickert, J. High accuracy optical flow estimation based on a theory for warping. *Lecture Notes in Computer Science* 3024 Part IV 25–36 (Springer, 2004).
- Keller, R. E. An experimental analysis of the role of bottle cells and the deep marginal zone in gastrulation of *Xenopus laevis*. *J. Exp. Zool.* **216**, 81–101 (1981).
- Rohani, N., Canty, L., Luu, O., Fagotto, F. & Winklbauer, R. EphrinB/EphB signaling controls embryonic germ layer separation by contact-induced cell detachment. *PLoS Biol.* **9**, e1000597 (2011).
- Tuft, P. H. The uptake and distribution of water in the embryo of *Xenopus laevis* (Daudin). *J. Exp. Biol.* **39**, 1–19 (1962).
- Deans, S. R. *The Radon Transform and Some of Its Applications* Ch. 6 (Wiley, 1983).
- Kashef, J. *et al.* Cadherin-11 regulates protrusive activity in *Xenopus* cranial neural crest cells upstream of Trio and the small GTPases. *Genes Dev.* **23**, 1393–1398 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to acknowledge discussions with J. Wittbrodt, H. Steinbeisser, R. Winklbauer and D. Moss. R. Keller and D. Shook helped us with interpreting the data. D. Wedlich and M. Köhl commented on the manuscript. Discussions with M. Köhl on data analysis are gratefully acknowledged. We also would like to thank T. van de Kamp and D. Karpov for their help visualising the set-up, as well as F. de Carlo and K. Fezza for allocating beamtime at 2-BM-B station and at 32-ID, respectively, of Advanced Photon Source, Argonne National Laboratory. Use of the Advanced Photon Source, an Office of Science User Facility operated for the US Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the US DOE under Contract No. DE-AC02-06CH11357. E. Becker, C. Huang, A. Merks and R. Langhe helped analysing blastopore radii. J.K.'s Young Investigator Group received financial support from the 'Concept for the Future' programme of Karlsruhe Institute of Technology within the framework of the German Excellence Initiative. This research partially was funded by the German Federal Ministry of Education and Research under grant numbers 05K12CK2 and 05K12VH1.

Author Contributions J.M., J.K. and R.H. conceived, organised and planned the experiments at beamlines 2-BM-B and 32-ID of Advanced Photon Source including feasibility studies, data management and testing/optimizing the set-up. V.A. established first contact between physicists and developmental biologists at Karlsruhe Institute of Technology to investigate *X. laevis* embryos. The physics part of the experiment was performed by J.M., X.X. and R.H., the biological parts by J.K. C.LaB. and M.S.P. helped with preparing samples. X.X. conducted heat-load measurements. R.H. investigated heat load theoretically. J.M. and A.E. developed data pre-processing routines. J.M. performed data pre-processing, phase retrieval and tomographic reconstruction. A.E. developed optical flow methods, computed velocity fields and performed image analysis. A.E. and J.M. worked on visualisations. A.E., J.M., V.A. and J.K. performed segmentation of the data. A.E., J.M. and R.H. analysed cell trajectories. J.M., A.E., T.B., X.X., J.K. and R.H. continuously discussed physics and biology problems. R.H. drafted the manuscript, A.E., J.M. and J.K. discussed and contributed text. C.LaB., M.S.P. and T.B. provided critical review. All authors contributed to the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.H. (ralf.hofmann2@kit.edu) or J.K. (jubin.kashef@kit.edu).

METHODS

Sample preparation. *In vitro* fertilisation, embryo culture and staging were carried out as described³⁰.

Detector system. The scintillator was a LuAG:Ce of 100- μm thickness. The camera system consisted of a pco.dimax CMOS (Cooke) with $2,016 \times 2,016$ pixels of size $11 \times 11 \mu\text{m}^2$ and a Mitutoyo long working distance $\times 5$ lens (corresponding to an effective pixel size of $\Delta x = 2.2 \mu\text{m}$ and $\Delta x = 1.1 \mu\text{m}$ at 2-BM-B and 32-ID, respectively).

Data acquisition, processing, and analysis. The data was mainly collected at the bending magnet beamline station 2-BM-B of Advanced Photon Source (Argonne National Laboratory) with a monochromatization level of $\Delta E/E = 10^{-2}$ and an X-ray energy of $E = 30 \text{ keV}$, a flux density at the sample of 10^{12} photons per mm^2 per second, a sample–detector distance of $z = 62 \text{ cm}$, $N = 1,200$ projections per tomogram, an exposure per projection of 15 ms, and a waiting time between tomograms of $\sim 10 \text{ min}$ when no irradiation of the embryo took place. Time-lapse series were terminated as soon as apoptotic cells or ectoderm ruptures could be observed in the radiographs. Following reconstruction, we concluded from the sequence of volumes, analysed in detail in the present work, that the associated embryo had been imaged *in vivo* for approximately two hours of development (Supplementary Video 3). To assure that *in vivo* imaging of gastrulation can be performed with another X-ray source, we also recorded a sequence at beamline 32-ID of Advanced Photon Source (Argonne National Laboratory) with a monochromatization level of $\Delta E/E = 10^{-4}$, an X-ray energy of $E = 34 \text{ keV}$, a flux density at the sample of about two times 10^{12} photons per mm^2 per second, a sample–detector distance of $z = 70 \text{ cm}$, $N = 834$ projections per tomogram, an exposure per projection of 20 ms and a waiting time between tomograms of $\sim 8 \text{ min}$ (Supplementary Video 4 and Supplementary Fig. 3).

Apart from dose reduction due to an adjustment of X-ray energy, our two-hour time-lapse series, starting at developmental stage 11.5, could be acquired by exploiting the following points: (1) applicability of one exposure per tomographic projection by a single-distance set-up (Fig. 1) owing to a negligible contribution of absorption to the intensity contrast²², (2) use of a large propagation distance, bounded from above by image blurring due to the size of the X-ray source, to provide optimal image contrast²⁴, and (3) tuning down of exposure time per projection to 15 ms, thus scraping the contrast limit for image analysis.

Raw images were flat- and dark-field corrected including removal of hot pixels and monochromator modulation. After phase retrieval^{22,24}, volumes were reconstructed from the *in vivo* data resulting in a 4D view of mid-gastrulation with micrometre-scale spatial resolution (Supplementary Videos 2 and 3). For both data pre-processing and phase retrieval, MATLAB routines were written. Tomographic reconstruction was performed with a filtered back-projection routine²⁹ inbuilt into PyHST (developed by European Synchrotron Radiation Facility and optimised at Karlsruhe Institute of Technology). We rendered cells, tissues and morphological features manually using Amira/Avizo and ImageJ/Fiji software^{31,32}.

Velocity fields were computed by an automated 3D optical flow algorithm (developed at Karlsruhe Institute of Technology) built to minimise the summed difference of grey values between displaced pixels in successive volumes. This technique relies on the formulation of a variational problem and determines the 3D dense displacement field as its unique solution^{25,33,34}. Cell trajectories were obtained by integrating the flow field, the locations of cellular nuclei serving as initial conditions. We checked the accuracies of the optical flow algorithm and of the integration routine by comparing the trajectories of manually tracked nuclei with the results of flow-field integration subject to the same initial positions of the nuclei. Compared to manually-tracked trajectories, the average error over four successive volumes is about $4 \mu\text{m}$ at an average distance of $58 \mu\text{m}$ travelled. Moreover, time-interpolated cell positions of four successive (uncorrupted) volumes between 52 min and 83 min were obtained by determining for each coordinate the coefficients of a cubic in terms of the data. To compute velocities, the according interpolated trajectories were differentiated with respect to time. With future data (longer time-lapse sequences) higher-order time derivatives should be computed of polynomials whose order is significantly smaller than the number of fit points. We estimate the maximum acceleration magnitude throughout mid-gastrulation as $0.2 \mu\text{m per min}^2$.

31. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671–675 (2012).
32. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676–682 (2012).
33. Horn, B. K. P. & Schunck, B. G. Determining optical flow. *Artif. Intell.* **17**, 185–203 (1981).
34. Baker, S. et al. A database and evaluation methodology for optical flow. *Int. J. Comput. Vis.* **92**, 1–31 (2011).

The shaping and functional consequences of the microRNA landscape in breast cancer

Heidi Dvinge^{1*†}, Anna Git^{1*}, Stefan Gräßl[†], Mali Salmon-Divon^{2,3}, Christina Curtis⁴, Andrea Sottoriva⁴, Yongjun Zhao^{5,6}, Martin Hirst^{5,6}, Javier Armisen⁷, Eric A. Miska⁷, Suet-Feung Chin¹, Elena Provenzano⁸, Gulisa Turashvili^{6,9}, Andrew Green¹⁰, Ian Ellis¹⁰, Sam Aparicio^{6,9} & Carlos Caldas^{1,8,11}

MicroRNAs (miRNAs) show differential expression across breast cancer subtypes, and have both oncogenic and tumour-suppressive roles^{1–6}. Here we report the miRNA expression profiles of 1,302 breast tumours with matching detailed clinical annotation, long-term follow-up and genomic and messenger RNA expression data⁷. This provides a comprehensive overview of the quantity, distribution and variation of the miRNA population and provides information on the extent to which genomic, transcriptional and post-transcriptional events contribute to miRNA expression architecture, suggesting an important role for post-transcriptional regulation. The key clinical parameters and cellular pathways related to the miRNA landscape are characterized, revealing context-dependent interactions, for example with regards to cell adhesion and Wnt signalling. Notably, only prognostic miRNA signatures derived from breast tumours devoid of somatic copy-number aberrations (CNA-devoid) are consistently prognostic across several other subtypes and can be validated in external cohorts. We then use a data-driven approach⁸ to seek the effects of miRNAs associated with differential co-expression of mRNAs, and find that miRNAs act as modulators of mRNA–mRNA interactions rather than as on–off molecular switches. We demonstrate such an important modulatory role for miRNAs in the biology of CNA-devoid breast cancers, a common subtype in which the immune response is prominent. These findings represent a new framework for studying the biology of miRNAs in human breast cancer.

Breast cancer is a clinically heterogeneous disease, with tumour subtypes defined either histopathologically (oestrogen receptor (ER)^{+/–}, human epidermal growth factor receptor (HER2)^{+/–}), using a gene-expression-based classifier (PAM50; Basal, HER2, Luminal A, Luminal B and Normal-like subtypes)⁹ or using the new integrative classification based on genomic and transcriptomic data (iClusters 1–10)⁷.

MicroRNAs (miRNAs) are short (20–23 nucleotides) regulatory RNAs that cause destabilization or translational repression of target mRNAs. The role of individual miRNAs in breast cancer¹ has mostly been investigated^{2–6} using small cohorts lacking matched genomic, mRNA or long-term survival data (Supplementary Fig. 1a, b). This study probes a cohort of breast tumours ($n = 1,302$) representing the heterogeneity observed in the clinic⁷ (Supplementary Fig. 1c and Supplementary Table 1), as well as adjacent normal breast tissue ($n = 116$) and a panel of breast cancer cell lines ($n = 28$).

Measurement of miRNA levels relative to non-miRNA array content demonstrates a global decrease in miRNA expression in tumours

compared to adjacent normal tissue, and a gradual decline with increased tumour grade and in breast cancer cell lines (Supplementary Fig. 2a). This is only partly explained by accompanying changes in miRNA-processing genes (Supplementary Fig. 2b). The 853 detectable miRNAs (Supplementary Table 2) partition into three main classes: 43% with very low variability and two blocks (22% and 35%) with globally opposite patterns of high/low expression (Supplementary Fig. 2c). We note in particular the steady expression of several miRNAs, which can serve as reliable reference genes in quantitative reverse-transcriptase PCR studies (Supplementary Fig. 2d).

There was no observed location bias for detected versus undetected miRNAs (Fig. 1a) and we found 226 miRNAs residing in regions affected by somatic CNAs in at least 10% of samples (Fig. 1a tracks 1–3; 636 miRNAs at 5%). Although miRNA genes are over-represented in copy-altered compared to copy-neutral regions ($P = 10^{-7}$; Fisher's exact test in bins of 22 base pairs), this enrichment mirrors that of coding genes (Supplementary Fig. 3a). In contrast to mRNAs (Fig. 1b and Supplementary Fig. 3b), few miRNAs have expression levels driven by CNAs (Fig. 1a track 4 and Supplementary Table 3), and this effect can be context dependent (for example, the uniformly CNA-driven miR-30b compared to ER-status-dependent miR-342; Supplementary Fig. 3c). We identified 133 miRNAs resident in minimal common regions (MCRs) with recurrent CNAs in at least 5% of the samples (Supplementary Table 4a). Examples include gains of the miR-17-92 oncogenic family¹⁰ and loss of the tumour-suppressive miR-31 (ref. 11). Twenty-four of the miRNAs, including seven putative new miRNAs, are in MCRs containing no other genes (Supplementary Table 4b). The frequency of miRNAs in MCRs was not systematically higher than that of protein-coding genes ($P = 0.74$ –1; Fisher's exact test in all samples, ER⁺ samples and ER[–] samples).

We used the correlation between miRNAs and their nearest protein-coding gene to evaluate the effect of gross changes to chromatin architecture on gene expression (Fig. 1c and Supplementary Fig. 3d). This surrogate measure of gross epigenetic effects is a key contributor (correlation >0.4) to variation in expression in less than 1% of miRNAs. Concordantly, the density of CpG islands around promoters of miRNA genes is lower than around mRNA promoters (Fig. 1d).

Only 49 out of 227 same-strand intragenic miRNAs exhibit high (>0.4) correlation with the host mRNA, suggesting a limited miRNA–host co-transcription (Fig. 1c). Consistently, co-regulated miRNA–host mRNA genes include the ER⁺ marker miR-342 (refs 3, 12) and the ER-regulated *EVL*¹³. Other miRNA–host relationships are subtype

¹Cancer Research UK Cambridge Institute and Department of Oncology, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. ²European Molecular Biology Laboratory–European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK. ³Department of Molecular Biology, Ariel University Center of Samaria, Ariel 40700, Israel. ⁴Department of Preventive Medicine, University of Southern California, Los Angeles, California 90033, USA. ⁵Genome Sciences Centre, BC Cancer Agency, 675 West 10th Avenue, Vancouver V5Z 1L3, Canada. ⁶Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver V6T 2B5, Canada. ⁷Wellcome Trust Cancer Research UK Gurdon Institute and Department of Biochemistry, University of Cambridge, The Henry Wellcome Building of Cancer and Developmental Biology, Cambridge CB2 1QN, UK. ⁸Cambridge Breast Unit, Addenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust and NIHR Cambridge Biomedical Research Centre, Cambridge CB2 2QQ, UK. ⁹Molecular Oncology, British Columbia Cancer Research Centre, Vancouver V5Z 1L3, Canada. ¹⁰Department of Histopathology, School of Molecular Medical Sciences, University of Nottingham, Nottingham NG5 1PB, UK. ¹¹Cambridge Experimental Cancer Medicine Centre, Cambridge CB2 0RE, UK. [†]Present addresses: Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA (H.D.); Department of Medicine, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0QQ, UK (S.G.).

*These authors contributed equally to this work.

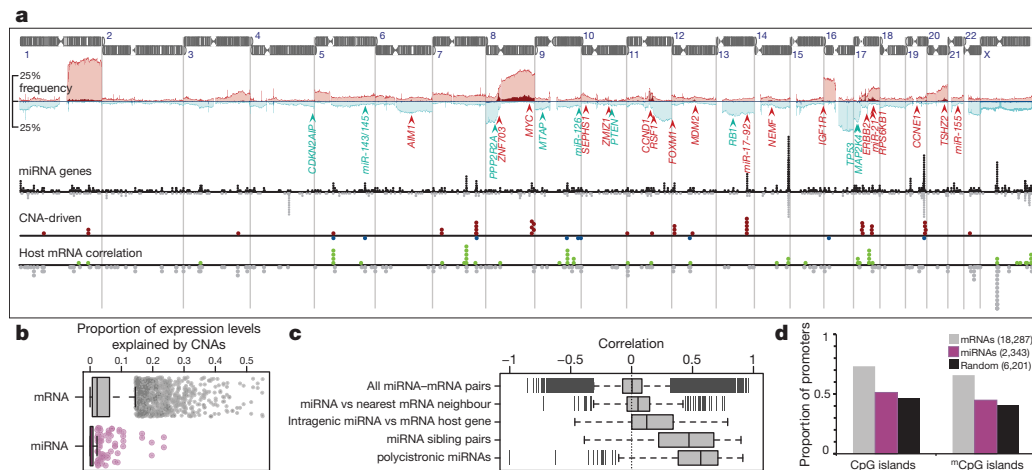


Figure 1 | Factors shaping the miRNA landscape across the breast cancer genome. **a**, miRNA position, detectability and association with CNAs and host gene transcription. Panels top to bottom: chromosome schematic; CNA frequency across cohort (red: amplification, >3 copies; pink: gain, 3 copies; light blue: heterozygous deletion, 1 copy; dark blue: homozygous deletion, 0 copies) and key breast cancer genes⁵ and miRNAs (red, oncogene; blue, tumour suppressor); location of miRNAs (black, detected; grey, non-detected); miRNAs significantly altered by CNAs (red, amplifications/gains; blue, heterozygous/homozygous deletions); miRNAs resident inside coding genes (green, correlated miRNA–host expression; grey, uncorrelated). miRNA counts are summarized across 100-kb bins. **b**, miRNA expression is

specific, and combine independent constitutive or regulated miRNA promoters with host co-transcribed contributions. For instance, miR-10a and miR-505 seem to be co-transcribed with their host genes *HOXB3* and *HOXB4* or *ATP11C* only in ER⁺ or ER[−] samples, respectively (Supplementary Fig. 3e).

Co-transcription of multiple pre-miRNAs from a polycistronic primary transcript is substantially more dominant, as seen in the high average correlation (0.49 ± 0.31) inside a 210-kilobase (kb) polycistron on chromosome 14q32 (ref. 14) (median consecutive distances: 1.2–111 kb), spanning 47 detectable miRNAs (Supplementary Fig. 3f). Putative polycistronic miRNAs can show dependent, independent or even opposite expression patterns (exemplified by miR-494, miR-134 and miR-370; Supplementary Fig. 3g), suggestive of separate promoters, or alternatively of a dominant post-transcriptional regulation of miRNA levels¹⁵. Evidence for the latter is clearly observed in the variable expression ratio for pairs of mature miRNAs arising from a single pre-miRNA (Supplementary Fig. 3h and Supplementary Table 5).

We established the relationship of the miRNA landscape with clinical, histological and molecular parameters⁷ using a generalized additive model (GAM) describing the expression pattern of 302 variable miRNAs (interquartile range >0.58, Fig. 2a) or 5,367 variable mRNAs (Supplementary Fig. 4a). Notably, all tested parameters explain more of the variation in mRNA expression than in miRNA expression. None of the clinical parameters stood out as a key covariate, and of the histopathological markers, ER had the largest effect, with a subset of miRNAs varying systematically among ER and PAM50 subtypes^{3,4} (Supplementary Table 6). Lymphocytic infiltration showed a very pronounced effect on five miRNAs with known functions in immune system regulation: miR-150 (ref. 16), miR-155 (ref. 17), miR-146a¹⁸, miR-142-3p and miR-142-5p¹⁹.

In contrast to the clinical covariates, molecular signatures (PAM50 or iClust) and individual mRNA variation were associated with a large proportion of the miRNA expression architecture (validated by comparing miRNA–mRNA GAM values in our data set with the corresponding values derived from the TCGA data set⁶, Supplementary Fig. 4b). Unsupervised clustering of the mRNA–miRNA GAM values (Fig. 2b and Supplementary Table 7) resulted in ten mRNA branches.

uncommonly copy-number driven. The proportion of mRNA or miRNA expression variation explained by underlying copy-number changes, estimated by a GAM. **c**, Contributors to miRNA expression. Pairwise correlations between: all miRNA–mRNA pairs; each miRNA and its nearest mRNA neighbour, including 91 miRNAs on opposite strand of mRNA, 227 miRNAs encompassed by the coding strand of mRNAs, 59 miRNA pairs derived from a single pre-miRNA and 42 miRNAs arising from putative polycistrons. **d**, miRNA promoters span fewer CpG islands than mRNA promoters. The proportion of mRNA, miRNAs or a random selection of >6,000 genomic loci whose nearest upstream promoter-proximal regulatory module is within 2 kb of predicted CpG or methylated CpG (mCpG) islands.

Branches 1 and 2 comprise unconnected miRNAs and mRNAs, and are enriched in ‘housekeeping’ mRNAs involved in transcription, translation and subcellular transport (false discovery rate <0.01; Supplementary Table 8). The highly interdependent miRNAs and mRNAs represented several key cellular processes associated with breast cancer biology and survival²⁰. Branch 6 was highly enriched in terms related to immune response and the five miRNAs most strongly associated with this branch are identical to those identified by lymphocytic infiltration (see above). Branches 3 and 5 were enriched in terms pertaining to regulatory components of signalling and development, extracellular matrix (ECM), cell adhesion and morphogenesis, interacting with known tumour-suppressive miRNA families such as miR-143/145 (ref. 21), miR-199/214 (ref. 22) and miR-127 (ref. 23). Branch 4 terms were almost exclusively related to the cell cycle, led by the anti-proliferative miR-139 (ref. 24). Branch 7 contained terms related to sex-hormone activity, reflecting the activity of oestrogen and progesterone in breast cancer, and led by the ER-dependent miR-342 (see above) and oncogenic miR-17-92 polycistron²⁵. Branches 8–10 exhibited a pattern of mRNA–miRNA associations akin to branches 3 and 6, suggesting that they are secondary to ECM- and cell-cycle-driven events. The mRNA–miRNA landscape is dominated by positive associations (Fig. 2c), indicating co-transcriptional modules. This includes pairs comprising miRNAs and their predicted or experimentally validated targets, suggesting that downregulation of target mRNA levels by miRNAs across the global breast tumour cohort is minor.

To evaluate context-specific mRNA–miRNA relationships we contrasted GAM values derived from ER⁺ and ER[−] cancers (Fig. 2d). Distinct interdependent relationships were evident, particularly in an ER-negative context. These were enriched in ECM/cell adhesion and immune response Gene Ontology terms and a mesenchymal stem cell gene set²⁶. In ER⁺ tumours we find components of the epithelial–mesenchymal-transition-related canonical Wnt pathway²⁷. For example, SFRP1 is linked to the tumour-suppressive miR-99a in ER⁺ but not ER[−] samples, whereas in the absence of the dominant oestrogen signal, miR-342 was tightly correlated with mRNAs of immune origin, consistent with its expression in leukocytes²⁸. The miRNA links of mRNAs pertaining to mammary gland epithelial proliferation were almost mutually exclusive between ER⁺ and ER[−] cancers.

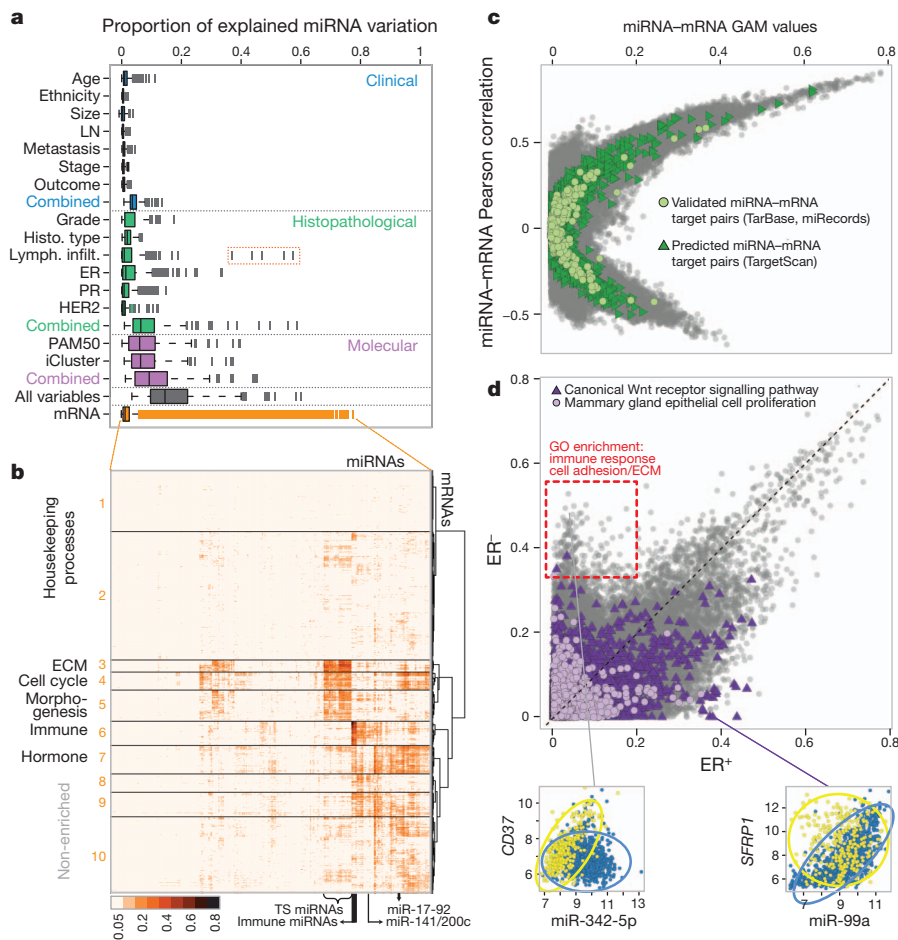


Figure 2 | Cellular processes connected to the miRNA landscape. **a**, miRNA variation versus sample meta-data and mRNA expression. GAM evaluation of the contribution of all variable mRNAs or established clinical, histopathological or molecular parameters to the observed variation in expression of 302 variable miRNAs (intensity 9.36 ± 2.24). miRNAs related to immune response (miR-150, miR-155, miR-142-5p, miR-142-3p and miR-146a) are boxed. LN, lymph-node positive; PR, progesterone receptor status. **b**, Groups of miRNAs are concerted by key cellular pathways. Unsupervised clustering of GAM values from all variable mRNA-miRNA combinations (Euclidean distance; Ward's minimum variance agglomeration). Gene Ontology terms significantly enriched in particular branches are listed (Supplementary Table 8). ECM, extracellular matrix; TS, tumour suppressor. **c**, The miRNA-mRNA landscape is dominated by positive associations. GAM values versus the directional Pearson correlation values for all mRNA-miRNA pairs or only pairs comprising miRNAs and their predicted or experimentally validated targets. **d**, miRNA-mRNA relations differ systematically in ER⁺ and ER⁻ samples. The GAM values calculated for ER⁺ and ER⁻ samples separately are contrasted, with mRNA-miRNA pairs pertaining to named Gene Ontology (GO) terms highlighted. Insets show expression of individual example pairs (generalized trends discussed in the text are circled approximately).

Individual prognostic miRNAs identified by a univariate Cox proportional hazards model mostly lose their significance in a multivariate model accounting for key clinical parameters (Supplementary Fig. 5a). Correcting the analysis for multiple testing leaves only a handful of miRNAs significant across the entire cohort (Fig. 3a) and their prognostic value rarely withstands validation in external data sets²⁻⁶ (Fig. 3b). Whereas miR-345 remained prognostic across the studies, miR-663 and miR-99a were varied and even prognostic of opposite outcomes, despite the absence of systematic differences, for example ER-driven differential expression (Supplementary Fig. 5b). This inconsistency can partly be attributed to differences in clinical cohort, as the prognostic utility of miRNAs was higher within ER⁺ tumours, and the Luminal B subtype (Fig. 3a). Notably, the prognostic value of miRNAs stands out in the genomically quiescent integrative cluster 4 subtype⁷, representing 19% (252) of our cohort. This observation is probably driven by the underlying molecular pathogenesis, as an artificial iClust4mimic subtype comprised of grade- and ER-matched tumours (Supplementary Table 1) failed to reproduce it.

A semi-supervised approach that uses principal component analysis of miRNAs with moderate individual significance ($P < 0.2$) in each subtype assigns weights to miRNAs and generates a combined signature²⁹. This first principal component (PC1) stratifies iClust4 patients into continuous prognostic groups, but not in Luminal B or ER⁺ subtypes (Fig. 3c). The iClust4 signature contains 41 miRNAs, mostly expressed at low levels (17 canonical miRNAs from the miRBase database (release 16), and 24 putative novel miRNAs; Supplementary Fig. 5c). We note that in general the putative novel miRNAs are more likely to be indicators of moderately worse prognosis (Supplementary Fig. 5d; single-sided t -test $P < 10^{-5}$), which may reflect their discovery by sequencing high-grade tumours and cell lines.

Restricting the iClust4 principal component analysis to canonical miRNAs retains a signature which is prognostic, albeit more weakly, within other subtypes as well as in external data sets, suggesting that the iClust4-derived signature is not context restricted (Fig. 3d and Supplementary Fig. 5e), except with biased cohort compositions (Supplementary Fig. 5f). Signatures derived from other subtypes are not consistently prognostic (exemplified by a significant but opposite-prognosis Luminal B signature; Supplementary Fig. 5g), suggesting that the iClust4 miRNA landscape is consistently linked with survival. Motivated by this observation, we proposed that miRNAs have a pathogenic role in these tumours, and took an integrative data-driven approach to seek miRNAs that act as master regulators⁸. We postulated that rather than acting as on-off switches of particular mRNAs, most miRNAs exert their effect by modulating the relationship between effector and target mRNAs, conceptually acting as co-repressors or co-activators and leading to differential co-expression of mRNAs, as exemplified by miR-301a (Fig. 4a).

We examined ~1,000 possible mRNA effectors and ~5,000 possible mRNA targets within subtypes with sufficient samples (>200) to allow statistically robust calculations⁸. Each target-effector pair was tested in subpopulations with high or low levels of individual miRNAs (potential modulators), avoiding combinations with passenger effects driven by a biased abundance of the tested mRNAs. Supplementary Fig. 6a, b summarizes the approach. The average number of targets per modulator varied fivefold (ER⁺, 17 ± 8 ; ER⁻, 42 ± 26 ; iClust4, 84 ± 39 ; iClust4mimic, 43 ± 20). ER⁺ tumours had the fewest effector-target pairs per modulator and hence the lowest miRNA modulatory activity (Fig. 4b), in contrast to the highly connected modulators and targets in iClust4 (Fig. 4c). Intermediate counts of modulators, effectors and targets were measured for ER⁻ tumours and iClust4mimic, suggesting

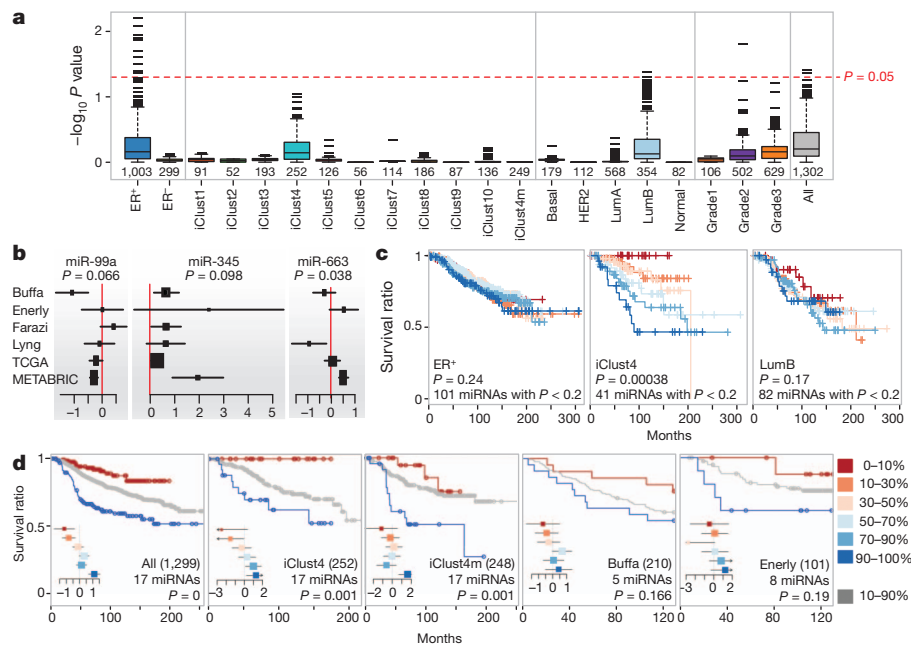


Figure 3 | miRNAs have an increased prognostic value in the genomically stable iClust4. **a**, Individual miRNAs are not robust prognostic factors in breast cancer. Survival prognosis by individual miRNAs, using a multivariate Cox proportional hazard model adjusted for age at diagnosis, tumour size, lymph-node status, grade and ER status and corrected for multiple testing. Numbers below boxes denote number of samples in each subtype. iClust4m, iClust4mimic. **b**, Example hazard ratios across multiple studies. Log₂ proportional hazard ratios and 95th confidence intervals across external data sets (Supplementary Fig. 1a) for miRNAs with adjusted *P* values <0.1 across our entire cohort. Size of squares are proportional to the weight assigned to each result. **c**, Compound miRNA signatures are more prognostic in iClust4. PC1

was calculated for subsets of miRNAs with an individual prognostic value of <0.2 in select subtypes, and used as a stratifier for Kaplan–Meier survival analysis. *P* values: multivariate Cox analyses. **d**, The iClust4 miRNA signature is prognostic in other subtypes and external cohorts. Kaplan–Meier survival analysis stratified by the PC1 based on 17 canonical miRNAs from iClust4; validated in external data sets with a balanced sample composition. Proportion of samples divided by PC1 is colour coded as in the key. Numbers of samples are in brackets. ‘miRNAs’ denotes the number of the 17 iClust4 miRNAs represented in the study. *P* values: univariate Cox model (not all stratifying variables were available for the external data sets). Detailed PC1-stratified log₂ proportional hazard ratios are shown in insets, cropped at (−3,2).

that at least some of the heterogeneity among these tumours may be related to altered miRNA activity.

All potential target mRNAs in iClust4 were ranked on the basis of a number of miRNAs modulating them, and tested for enrichment of biological processes. A substantial proportion of iClust4 samples exhibit extensive lymphocytic infiltration and a concerted immune

response network⁷, so we focused on descendants of ‘immune response’ (Gene Ontology code: 0006955). In the resulting modulation network (Fig. 4c and Supplementary Fig. 7a) we note miR-574-5p and members of the miR-27 and miR-29 families, previously linked to Toll-like-receptor-mediated induction of immune response³⁰. ‘Signal transduction inhibiting the immune response’ forms a densely connected

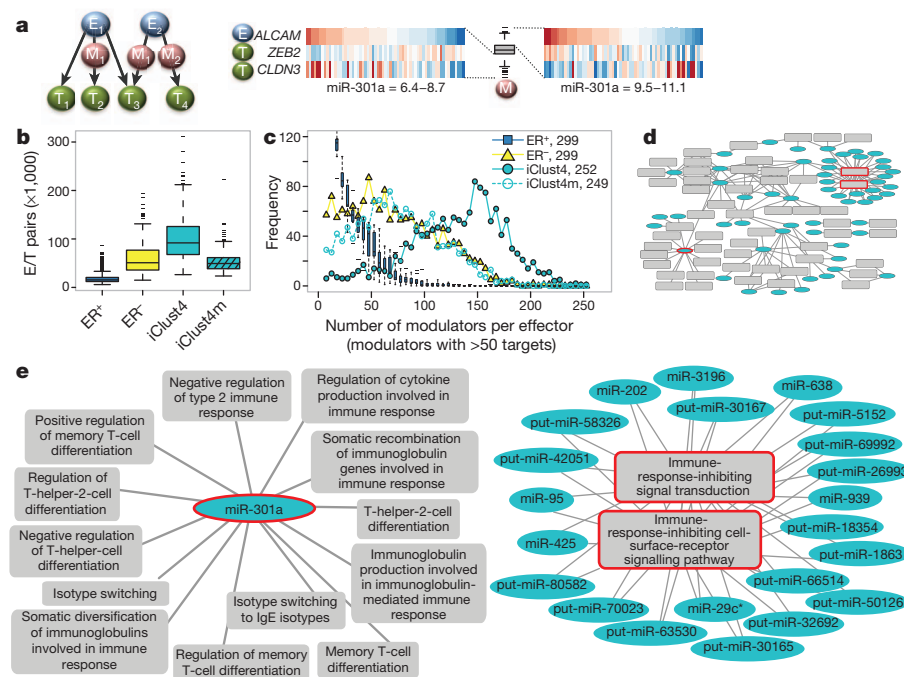


Figure 4 | miRNAs have a significant modulatory role in iClust4. **a**, Example of effector–modulator–target relationship. Left, effector (E)–modulator (M)–target (T) schematic. Right, the interaction between effector mRNA *ALCAM* and its two target mRNAs, *ZEB3* and *CLDN3*, differs in the subsets of samples with high or low expression of miR-301a, classifying the miRNA as a conceptual modulator. **b**, miRNAs have an increased modulator activity in iClust4. The number of effector–target pairs for each potential miRNA modulator across tumour subtypes. **c**, iClust4 has an increased number of highly connected modulators and effectors. Distribution of the number of highly connected modulators (>50 targets; arbitrary cutoff) per effector. For ER⁺ the boxes show the distribution across ten random samplings of 299 ER⁺ tumours, used as a contrast to the 299 ER[−] tumours. **d**, **e**, Network of immune-response-related Gene Ontology terms enriched among iClust4 targets. miRNAs (blue) modulate Gene Ontology terms descendants of ‘immune response’ (grey) in iClust4 (**d**). Highest connected miRNA and Gene Ontology term hubs are magnified (**e**). put-miR, putative miRNA.

hub (Fig. 4d), proffering the cumulative effect of multiple modulating miRNAs on immune response genes as an important mechanism in the iClust4 subtype, but not in the artificial iClust4mimic control (Supplementary Fig. 7b). miR-301a forms another hub, whose targets are enriched for T-cell-related processes unaffected by the remaining complement of miRNAs (Fig. 4d). Other mRNA relationships altered in miR-301a high/low tumours offer a link between immune and epithelial–mesenchymal transition pathways, illustrated by ALCAM, ZEB2 and CLDN3 (Fig. 4a). None of the miRNAs linked to lymphocytic infiltration (above) are associated with the enriched modulated biological processes, and because their expression levels are similar within the subtypes considered here, they do not drive the iClust4 immune network (Supplementary Fig. 7c).

In summary, we present a systems-level analysis of miRNA expression architecture in a large number of human breast tumours, by analysing miRNA levels and integrating them with matched mRNA expression and DNA copy number. This panoramic view shows the details of individual miRNAs that are expressed as classical oncogenes or tumour-suppressor genes, but mostly reveals that at the whole-genome level miRNAs behave more consistently as fine-tuners/modulators of gene expression. This is especially evident in the common subtype of breast cancer, in which copy-number events have no major role and miRNAs seem to modulate the immune response that characterizes these tumours.

METHODS SUMMARY

Tumour samples with linked pseudo-anonymized clinical data were obtained with appropriate ethical approval from their respective tumour banks. Quantification of known and putative miRNAs was performed using Agilent microarrays. The effects of CNAs and co-transcription on miRNA levels were measured using non-parametric and parametric models. The global interplay between miRNAs and mRNAs was analysed using a GAM, and the modulating effect of miRNAs was assessed using in-house scripts inspired by the MINDy algorithm⁸. Long-term patient outcomes were analysed using Cox proportional hazard models and Kaplan–Meier survival curves. Validation was performed with independent data sets^{2–6} generated on orthogonal platforms.

Received 2 August 2012; accepted 20 March 2013.

Published online 5 May 2013.

- Le Quesne, J. & Caldas, C. Micro-RNAs and breast cancer. *Mol. Oncol.* **4**, 230–241 (2010).
- Buffa, F. M. *et al.* microRNA-associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer. *Cancer Res.* **71**, 5635–5645 (2011).
- Enerly, E. *et al.* miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. *PLoS ONE* **6**, e16915 (2011).
- Farazi, T. A. *et al.* MicroRNA sequence and expression analysis in breast tumors by deep sequencing. *Cancer Res.* **71**, 4443–4453 (2011).
- Lyng, M. B. *et al.* Global microRNA expression profiling of high-risk ER⁺ breast cancers from patients receiving adjuvant tamoxifen mono-therapy: a DBCG study. *PLoS ONE* **7**, e36170 (2012).
- The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
- Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
- Wang, K. *et al.* Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. *Nature Biotechnol.* **27**, 829–837 (2009).
- Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
- Olive, V., Jiang, I. & He, L. mir-17–92, a cluster of miRNAs in the midst of the cancer network. *Int. J. Biochem. Cell Biol.* **42**, 1348–1354 (2010).
- Valastyan, S. & Weinberg, R. A. miR-31: a crucial overseer of tumor metastasis and other emerging roles. *Cell Cycle* **9**, 2124–2129 (2010).
- Blenkiron, C. *et al.* MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biol.* **8**, R214 (2007).
- Luqmani, Y. A., Al Azmi, A., Al Bader, M., Abraham, G. & El Zawahri, M. Modification of gene expression induced by siRNA targeting of estrogen receptor α in MCF7 human breast cancer cells. *Int. J. Oncol.* **34**, 231–242 (2009).
- Seitz, H. *et al.* A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome Res.* **14**, 1741–1748 (2004).
- Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nature Rev. Genet.* **11**, 597–610 (2010).
- Bezman, N. A., Chakraborty, T., Bender, T. & Lanier, L. L. miR-150 regulates the development of NK and iNKT cells. *J. Exp. Med.* **208**, 2717–2731 (2011).
- Faraoni, I., Antonetti, F. R., Cardone, J. & Bonmassar, E. miR-155 gene: a typical multifunctional microRNA. *Biochim. Biophys. Acta* **1792**, 497–505 (2009).
- Xu, W. D., Lu, M. M., Pan, H. F. & Ye, D. Q. Association of microRNA-146a with Autoimmune Diseases. *Inflammation* **35**, 1525–1529 (2012).
- Andreopoulos, B. & Anastassiou, D. Integrated analysis reveals hsa-miR-142 as a representative of a lymphocyte-specific gene expression and methylation signature. *Cancer Inform.* **11**, 61–75 (2012).
- Teschendorff, A. E., Miremadi, A., Pinder, S. E., Ellis, I. O. & Caldas, C. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. *Genome Biol.* **8**, R157 (2007).
- Peng, X. *et al.* Identification of miRs-143 and -145 that is associated with bone metastasis of prostate cancer and involved in the regulation of EMT. *PLoS ONE* **6**, e20341 (2011).
- Li, B. *et al.* Down-regulation of miR-214 contributes to intrahepatic cholangiocarcinoma metastasis by targeting Twist. *FEBS J.* **279**, 2393–2398 (2012).
- Git, A. *et al.* PMC42, a breast progenitor cancer cell line, has normal-like mRNA and microRNA transcriptomes. *Breast Cancer Res.* **10**, R54 (2008).
- Guo, H., Hu, X., Ge, S., Qian, G. & Zhang, J. Regulation of RAP1B by miR-139 suppresses human colorectal carcinoma cell proliferation. *Int. J. Biochem. Cell Biol.* **44**, 1465–1472 (2012).
- Castellano, L. *et al.* The estrogen receptor- α -induced microRNA signature regulates itself and its transcriptional response. *Proc. Natl Acad. Sci. USA* **106**, 15732–15737 (2009).
- Boquest, A. C. *et al.* Isolation and transcription profiling of purified uncultured human stromal stem cells: alteration of gene expression after *in vitro* cell culture. *Mol. Biol. Cell* **16**, 1131–1141 (2005).
- Scheel, C. *et al.* Paracrine and autocrine signals induce and maintain mesenchymal and stem cell states in the breast. *Cell* **145**, 926–940 (2011).
- Schmidt, W. M., Spiel, A. O., Jilma, B., Wolz, M. & Muller, M. *In vivo* profile of the human leukocyte microRNA response to endotoxemia. *Biochem. Biophys. Res. Commun.* **380**, 437–441 (2009).
- Bair, E. & Tibshirani, R. Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol.* **2**, e108 (2004).
- Fabbri, M. *et al.* MicroRNAs bind to Toll-like receptors to induce prometastatic inflammatory response. *Proc. Natl Acad. Sci. USA* **109**, E2110–E2116 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements The study was funded by Cancer Research UK and the British Columbia Cancer Foundation. The authors also acknowledge the support of the University of Cambridge, Hutchinson Whampoa, the NIHR Cambridge Biomedical Research Centre, and the Cambridge Experimental Cancer Medicine Centre. We thank S. McGuire for help in sample management; S. Fulmer-Smentek and T. Hill for help with array design; L. Goldstein for initial processing of sequencing data; O. Rueda for statistical advice; and O. Rueda, J. Carroll and R. Ali for reading of the manuscript. We are very grateful to the patients who donated tissue and associated pseudo-anonymized clinical data.

Author Contributions H.D. and A.Git led the analysis and drafted the manuscript with C.Caldas; A.Git, S.G., S.A. and C.Caldas designed and coordinated the study; A.Git carried out all microarray and quantitative reverse transcriptase PCR laboratory work. Sequencing data were provided by Y.Z., M.H., J.A., E.A.M. and S.A. and analysed by M.S.-D., who also analysed external epigenetic data; S.G. designed custom microarray probes and contributed to array pre-processing; C.Curtis and A.S. processed external CNA data. S.-F.C., E.P., A.Green, I.E., G.T., S.A. and C.Caldas coordinated collection and processing of clinical material and associated clinical and histopathological information. S.A. and C.Caldas are joint senior authors and project co-leaders.

Author Information The raw non-coding RNA microarray data is available through the European Genome-Phenome Archive (<http://www.ebi.ac.uk/ega/>), which is hosted by the EBI, under accession number EGAS00000000122. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.Git (anna.git@cruk.cam.ac.uk), S.A. (saparicio@bccrc.ca) or C.Caldas (carlos.caldas@cruk.cam.ac.uk).

EGFR modulates microRNA maturation in response to hypoxia through phosphorylation of AGO2

Jia Shen^{1,2}, Weiya Xia¹, Yekaterina B. Khotskaya¹, Longfei Huo¹, Kotaro Nakanishi³, Seung-Oe Lim¹, Yi Du^{1,2}, Yan Wang¹, Wei-Chao Chang^{4,5}, Chung-Hsuan Chen⁵, Jennifer L. Hsu^{1,4,6}, Yun Wu⁷, Yung Carmen Lam¹, Brian P. James⁸, Xiuping Liu⁸, Chang-Gong Liu⁸, Dinshaw J. Patel³ & Mien-Chie Hung^{1,2,4,6}

MicroRNAs (miRNAs) are generated by two-step processing to yield small RNAs that negatively regulate target gene expression at the post-transcriptional level¹. Deregulation of miRNAs has been linked to diverse pathological processes, including cancer^{2,3}. Recent studies have also implicated miRNAs in the regulation of cellular response to a spectrum of stresses⁴, such as hypoxia, which is frequently encountered in the poorly angiogenic core of a solid tumour⁵. However, the upstream regulators of miRNA biogenesis machineries remain obscure, raising the question of how tumour cells efficiently coordinate and impose specificity on miRNA expression and function in response to stresses. Here we show that epidermal growth factor receptor (EGFR), which is the product of a well-characterized oncogene in human cancers, suppresses the maturation of specific tumour-suppressor-like miRNAs in response to hypoxic stress through phosphorylation of argonaute 2 (AGO2) at Tyr 393. The association between EGFR and AGO2 is enhanced by hypoxia, leading to elevated AGO2-Y393 phosphorylation, which in turn reduces the binding of Dicer to AGO2 and inhibits miRNA processing from precursor miRNAs to mature miRNAs. We also identify a long-loop structure in precursor miRNAs as a critical regulatory element in phospho-Y393-AGO2-mediated miRNA maturation. Furthermore, AGO2-Y393 phosphorylation mediates EGFR-enhanced cell survival and invasiveness under hypoxia, and correlates with poorer overall survival in breast cancer patients. Our study reveals a previously unrecognized function of EGFR in miRNA maturation and demonstrates how EGFR is likely to function as a regulator of AGO2 through novel post-translational modification. These findings suggest that modulation of miRNA biogenesis is important for stress response in tumour cells and has potential clinical implications.

Activated EGFR contained in intracellular vesicles is capable of activating intracellular signalling pathways before lysosomal degradation⁶. Importantly, proteins associating with internalized EGFR probably differ from those transducing signalling at the plasma membrane⁷, suggesting a higher degree of signalling complexity that is not well characterized. We identified AGO2 as a novel EGFR-interacting protein by mass spectrometric analysis (Supplementary Fig. 1), and validated their association by co-immunoprecipitation and pull-down assays (Supplementary Fig. 2). The juxtamembrane and kinase domain of EGFR is essential for binding with AGO2 at the amino-terminal region.

Human AGO2 was initially reported as a membrane-associated cytoplasmic protein⁸ and is the catalytic centre of RNA-induced silencing complex⁹ (RISC). AGO2 also associates with Dicer and TRBP (HIV-1 transactivating response RNA-binding protein, also known as TARBP2) to form the RISC-loading complex, which is involved in the second step of miRNA processing from precursor to mature miRNAs^{10,11}. To investigate the physiological role of EGFR-AGO2 interaction, we screened different upstream EGFR-activating stimuli, including ligands and stresses^{12–15}, in

HTC-1080 stable clone expressing split-half-YFP-fused EGFR and AGO2 (EGFR with the N-terminal domain of YFP fused to the C terminus and AGO2 with C-terminal domain of YFP fused to the C terminus; Supplementary Fig. 3a; YFP, yellow fluorescent protein), in which the YFP fluorescence can be reconstituted only on protein–protein association¹⁶ (Fig. 1a). Of the four different types of stimuli, hypoxic stress induced the strongest level of YFP fluorescence (Supplementary Figs 3b, c and 4), with distinct foci formed in cytoplasm (Fig. 1b), suggesting that internalized EGFR interacts with AGO2 in aggregates. Dynamic EGFR-AGO2 association was further validated in HeLa cells and various cancer cell lines by co-immunoprecipitation (Fig. 1c and Supplementary Fig. 5) and co-localization assays (Supplementary Figs 6 and 7), and was found to be RNase resistant (Supplementary Fig. 8), indicating that EGFR and AGO2 are direct physical interacting partners *in vivo*.

Hypoxia is known to upregulate EGFR¹⁴ and prolong its activation through retention in endocytic trafficking¹³. Indeed, hypoxia enhanced EGFR expression in late endosomes (multivesicular bodies; Supplementary Fig. 9, F2–F4), where it co-localized with AGO2 (Supplementary Fig. 10 and Fig. 1d) and co-fractionated (Supplementary Fig. 9) with RISC components (AGO2, DCP1A and GW182) as well as the RISC-loading complex (Dicer, TRBP and AGO2). Silencing GRB2, a key modulator of EGFR endocytosis¹⁷, diminished EGFR-AGO2 interaction (Supplementary Fig. 11), highlighting the importance of internalization. Furthermore, inhibition or silencing of hypoxia-inducible transcriptional factors¹⁸ (HIF1 α (HIF1A) and HIF2 α (EPAS1)) reduced EGFR-AGO2 association (Supplementary Fig. 12a and Fig. 1e) and co-localization (Supplementary Figs 12b and 13) under hypoxia, and stabilization of HIF1 α and HIF2 α (HIF1/2 α) triggered the interaction of EGFR and AGO2 under normoxia (Supplementary Fig. 14). Nonetheless, EGFR-AGO2 association in RCC4 cells (endogenous VHL null with constitutively expressed HIF1/2 α) was further strengthened by hypoxia regardless of the exogenous expression of wild-type VHL (Supplementary Fig. 15). These results indicate that stable expression of HIF1/2 α is sufficient to trigger EGFR-AGO2 interaction that is further enhanced by hypoxia, probably through an HIF1/2 α -independent mechanism.

To assess the functional importance of EGFR-AGO2 interaction in miRNA regulation, we profiled RNA expression in HeLa stable clones expressing scrambled control or short hairpin RNA (shRNA) against EGFR under normoxia or hypoxia by RNA deep sequencing (Supplementary Figs 16a and 17). Hierarchical clustering analysis of relative expression of precursor and mature miRNAs (scrambled versus shRNA against EGFR; Supplementary Fig. 16b) identified one distinct cluster of miRNAs under hypoxia (Fig. 2a, dashed box). In the presence of EGFR, the level of precursor miRNAs increased with a concomitant decrease in the expression of mature miRNAs. However, the maturation of this cluster of miRNAs was not significantly altered by EGFR under normoxia (Fig. 2a), implying that their processing from precursor to

¹Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030, USA. ²The University of Texas Graduate School of Biomedical Sciences at Houston, Houston, Texas 77030, USA. ³Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ⁴Center for Molecular Medicine and Graduate Institute of Cancer Biology, China Medical University, Taichung 402, Taiwan. ⁵Genomics Research Center, Academia Sinica, Nankang, Taipei 105, Taiwan. ⁶Asia University, Taichung 413, Taiwan. ⁷Department of Pathology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030, USA. ⁸Department of Experimental Therapeutics, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030, USA.

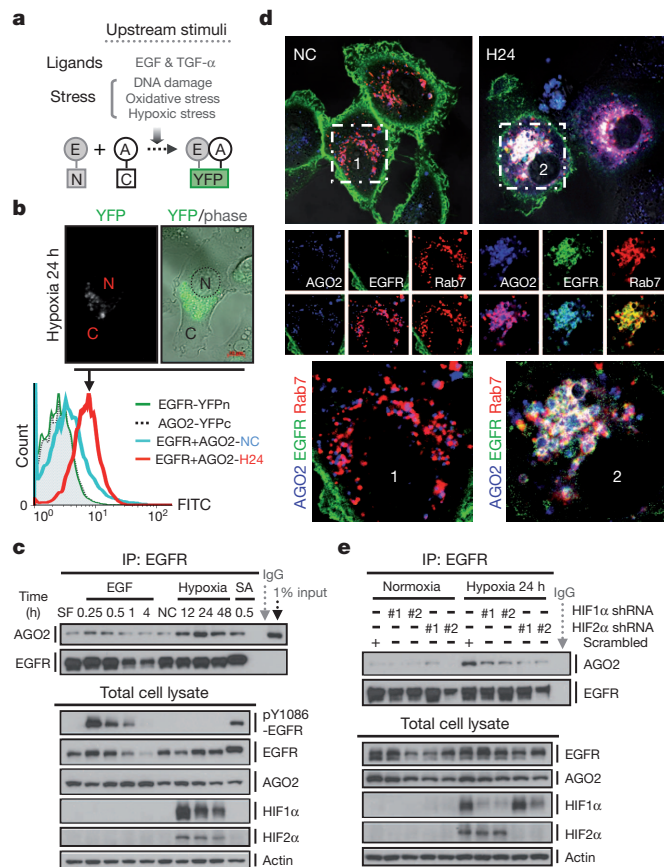


Figure 1 | EGFR interacts with AGO2 in response to hypoxia. **a**, Split-half YFP-fused EGFR and AGO2 were stably expressed in HTC-1080 cells to screen for upstream stimuli that might trigger EGFR-AGO2 interaction. E, EGFR; A, AGO2. **b**, Top, representative live-cell image. N, nuclear; C, cytoplasmic. Bottom, fluorescence-activated cell sorting (FACS) analysis of HTC-1080 stable transfectants as indicated. FITC, fluorescein isothiocyanate. **c**, Immunoprecipitation and western blot analysis of HeLa cells in response to different stimuli. EGF, 20 ng ml⁻¹; SA (sodium arsenite), 500 μ M; hypoxia, 1% O₂. **d**, Confocal microscopy analysis of live HeLa cells as indicated. Rab7, a marker for late endosomes. Regions 1 and 2 at top are shown in magnified view at bottom. NC, normoxia; H24, hypoxia 24 h. **e**, Immunoprecipitation and western blot analysis of HeLa stable transfectants expressing HIF1/2 α shRNAs.

mature miRNAs was negatively regulated by EGFR specifically in response to hypoxia. We defined this subcluster of miRNAs as mHESM (miRNAs regulated by hypoxia-dependent EGFR-suppressed maturation). We then pooled mHESM and narrowed down the candidates on the basis of their absolute mature miRNA expression affected by EGFR knockdown. A majority of the top-scoring mHESM turned out to have tumour suppressor characteristics^{3,4,19–21} (Fig. 2a).

To determine the functional relevance of mHESM, messenger RNAs (mRNAs) regulated by EGFR were sorted and overlapped with the predicted mRNA targets of top-scoring mHESM, revealing 439 mRNAs (Supplementary Data) that are regulated by EGFR and also targeted by top-scoring mHESM (Fig. 2b). In response to hypoxia, EGFR reduced the production of mHESM but enhanced the expression of corresponding mRNA targets (Fig. 2b), which is evidence of the importance of EGFR-modulated miRNA maturation. The inhibitory role of EGFR in miRNA maturation in response to hypoxia was further validated in HeLa stable transfectants expressing scrambled control or shRNAs targeting EGFR (Supplementary Fig. 19) by quantitative PCR. Moreover, induction of wild-type but not kinase-dead EGFR in HeLa Tet-Off-inducible stable clones (Supplementary Fig. 20a) inhibited the maturation of top-scoring mHESM in response to hypoxia (Supplementary

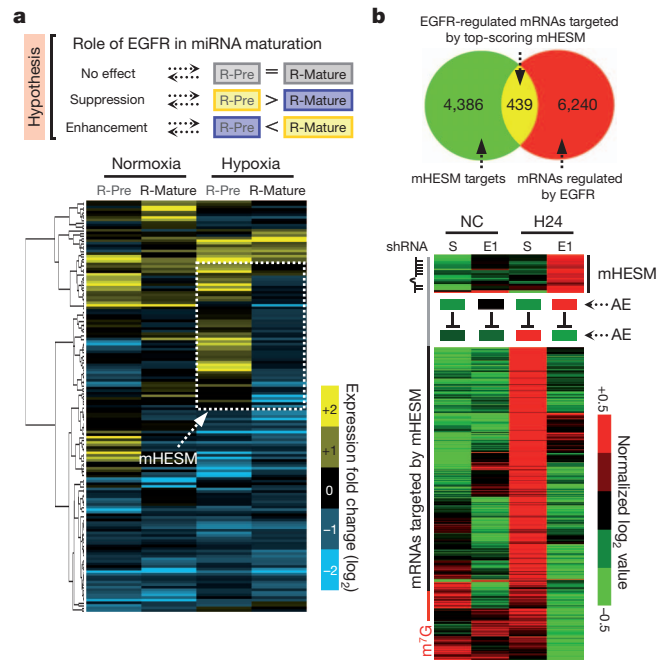


Figure 2 | EGFR modulates miRNA maturation in response to hypoxia. **a**, Top, proposed role of EGFR in miRNA maturation. Bottom, hierarchical clustering analysis of R-Pre (log₂[precursor miR (scrambled)] – log₂[precursor miR (EGFR shRNA-E1)]) and R-Mature (log₂[mature miR (scrambled)] – log₂[mature miR (EGFR shRNA-E1)]) identified one distinct cluster of miRNAs whose maturation was suppressed by EGFR under hypoxia. We define this subcluster as mHESM. **b**, Top, Venn diagram highlighting the mRNAs that are regulated by EGFR and likely to be targeted by top-scoring mHESM (those for which R-Pre – R-Mature \geq 0.8 and R-Mature \leq –0.4). Bottom, EGFR-mediated suppression of top-scoring mHESM concurrent with the upregulation of targeting mRNAs in response to hypoxia. AE, average expression.

Figs 20b,c), suggesting that EGFR kinase activity is important for EGFR-suppressed miRNA maturation.

To investigate whether AGO2 is a phosphorylation substrate of EGFR, we purified FLAG-tagged AGO2 co-expressed with EGFR and identified one Tyr phosphorylation site (Supplementary Fig. 21) at a highly conserved residue, Tyr 393 (Supplementary Fig. 22), in AGO2. Results from *in vitro* kinase assay (Fig. 3a) further demonstrated AGO2-Y393 to be a direct phosphorylation site targeted by EGFR. Mutational analysis suggested that phospho-Y393-AGO2 exists *in vivo* (Supplementary Figs 23 and 24), and this was validated in HeLa Tet-Off-inducible AGO2 stable clones (Supplementary Fig. 25a and Fig. 3b) using the polyclonal antibody (p-Y393-AGO2) we generated (Supplementary Fig. 26). Notably, hypoxia enhanced AGO2-Y393 phosphorylation (Fig. 3b), which in turn reduced the association of AGO2 with Dicer and TRBP (Fig. 3b and Supplementary Figs 23 and 24), suggesting that EGFR is a novel upstream regulator of the RISC-loading complex.

To gain more insight into phospho-Y393-AGO2, we analysed the crystal structure of AGO2 (refs 22 and 23) and found that the side chain of Tyr 393 protrudes with an exterior orientation towards a cavity between the N domain (an interaction surface for EGFR) and the linker L2 (a linker domain between PAZ and MID) (Fig. 3c). Tyr 393 is exposed to solvent and is some distance from both the guide RNA-binding channel and the PIWI box, a Dicer-binding region of AGO2 (ref. 24). Given that Dicer is a large protein, it is conceivable that Dicer could still interact with both the Dicer-specific PIWI box and Tyr 393 owing to their location on the same surface of AGO2 (Fig. 3c). If so, phosphorylation of Tyr 393 could inhibit this interaction as previously observed (Fig. 3b and Supplementary Figs 23 and 24).

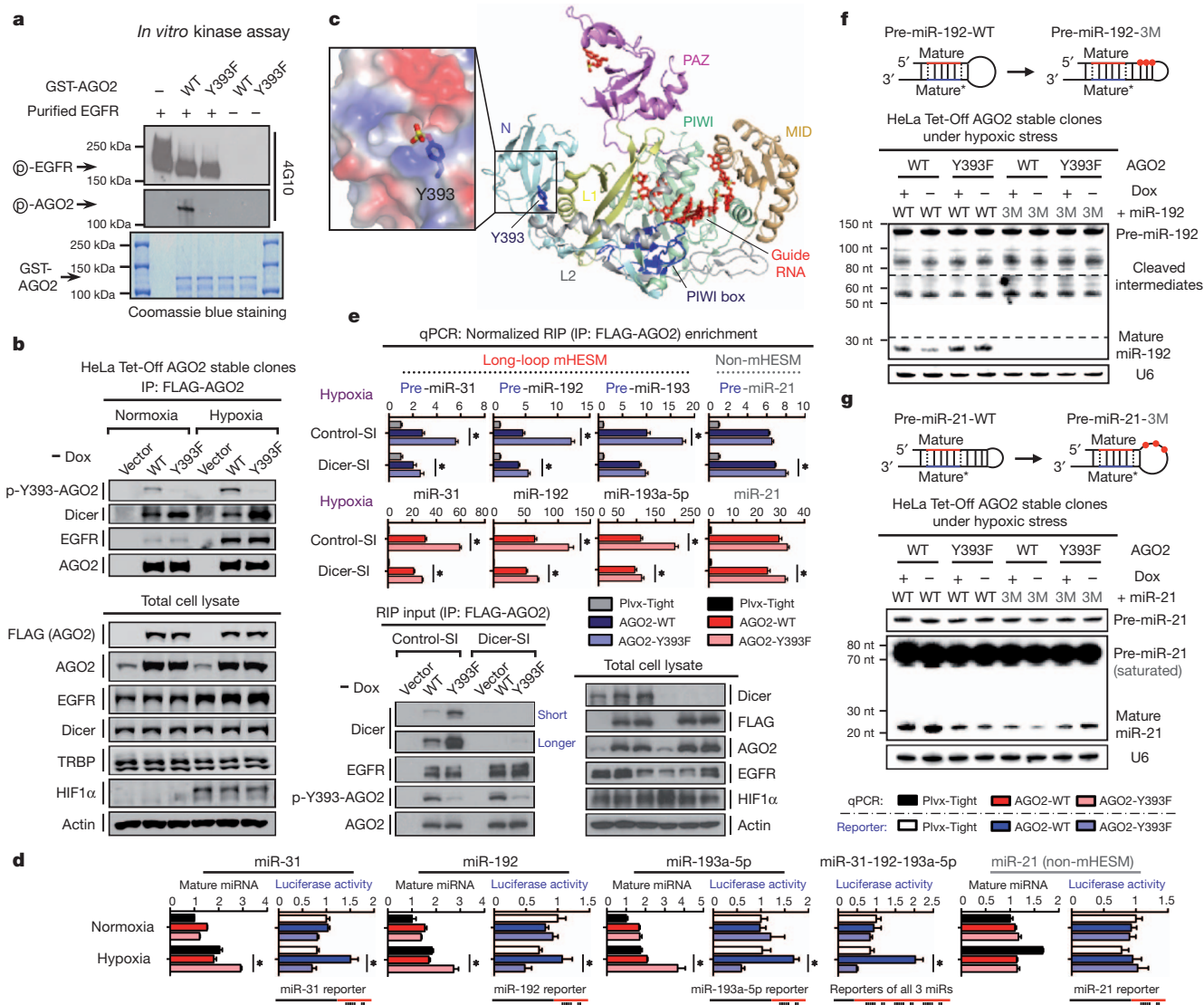


Figure 3 | EGFR phosphorylates AGO2 at Tyr 393 to suppress the maturation of long-loop mHESM in response to hypoxia. **a**, *In vitro* kinase assay detected by 4G10 antibody. **b**, Immunoprecipitation and western blot analysis of HeLa Tet-Off-inducible AGO2 stable clones. **c**, Relative position of Tyr 393 in the structure of human AGO2 with bound guide RNA^{22,23}. The N (cyan), linker L1 (yellow), PAZ (violet), linker L2 (grey), MID (orange) and PIWI (green) domains are shown in ribbon representation, and the co-purifying guide RNA (red) is depicted in stick representation. Tyr 393 and the PIWI box are highlighted in blue. The expanded boxed segment highlights the

electrostatic surface around Tyr 393, including a modelled phosphate attached to Tyr 393. **d**, Normalized expression of miRNA and corresponding luciferase activity of miR reporters (*n* = 4). **e**, RNA-binding protein immunoprecipitation (RIP) enrichment of precursor and mature miRNAs (*n* = 3). **f**, **g**, Northern blot analysis as indicated. Dashed lines indicate different exposure times. RNA integrity was examined by ethidium bromide staining (Supplementary Figs 32 and 33). Data represent mean \pm s.d.; **P* < 0.05, *t*-test.

The recruitment of AGO2 to Dicer is critical for loading miRNA precursors onto the RISC-loading complex²⁵ and facilitating miRNA maturation^{11,25} from precursor to mature miRNAs. We therefore investigated whether AGO2-Y393 phosphorylation has a role in EGFR-suppressed miRNA maturation in response to hypoxia. Compared with an AGO2-Y393F mutant, induction of wild-type AGO2 (AGO2-WT) significantly reduced the expression of most mHESM but not those that do not belong to the mHESM cluster (defined as non-mHESM) in response to hypoxia (Supplementary Fig. 25b). Structural analysis of miRNA precursors determined that a majority of mHESM regulated by p-Y393-AGO2 contained a long-loop structure, which is not present in non-mHESM (Supplementary Fig. 25b). Notably, mHESM that were not significantly affected by AGO2-Y393 phosphorylation also had short-loop structures in their precursors, similar to what we found in non-mHESM (Supplementary Fig. 25b). This suggests that long-loop structure is important in regulation

specificity. Similar expression patterns of mature miRNAs were observed in other paired stable clones (Supplementary Figs 27 and 28).

Silencing endogenous EGFR significantly diminished the expression difference of long-loop mHESM present in miR-31, miR-192 and miR-193a-5p between AGO2-WT and AGO2-Y393F mutant cells (Supplementary Fig. 29). The levels of their primary transcripts were reduced by EGFR knockdown under hypoxia but were similar in AGO2-WT and AGO2-Y393F cells (Supplementary Fig. 29b). This is evidence that EGFR is a tyrosine kinase that mediates phospho-Y393-AGO2-suppressed miRNA maturation under hypoxia. Moreover, decreased expression of long-loop mHESM as shown in miR-31, miR-192 and miR-193a-5p under hypoxia resulted in the de-repression of miRNA targets (Fig. 3d) as measured by miR-reporter luciferase activity. In contrast, the expression of miR-21 (non-mHESM) and the repression of its target were not significantly affected by AGO2-Y393 phosphorylation

(Fig. 3d). These results underline the functional importance of p-Y393-AGO2-mediated suppression of long-loop mHESM under hypoxia.

The long-loop structure in miRNA precursors is a known characteristic of Dicer's preference in substrate recognition²⁶. Reduction in the Dicer-AGO2 interaction resulted in less loading of the precursors of miR-31, miR-192 and miR-193a-5p (long-loop mHESM), but not that of miR-21 (short-loop non-mHESM), onto p-Y393-AGO2 under hypoxia (Supplementary Fig. 30). To examine the functional relevance of decreased Dicer-phospho-AGO2 association, we knocked down Dicer in HeLa Tet-Off-inducible AGO2 stable clones (Supplementary Fig. 31a, b and Fig. 3e) and found that the differences between AGO2-WT and AGO2-Y393F in miRNA precursor loading (Fig. 3e) and mature miRNA expression (Supplementary Fig. 31c) were significantly diminished. These results suggest that the maturation of long-loop mHESM is suppressed by AGO2-Y393 phosphorylation through Dicer. Moreover, AGO2-Y393F was capable of loading more mature mHESM (Fig. 3e), which is consistent with its enhanced RISC activity as indicated by luciferase reporter assay (Fig. 3d). However, the mature miRNA loading difference between AGO2-WT and AGO2-Y393F was Dicer dependent (Fig. 3e), and similar to what we observed in mature miRNA expression (Supplementary Fig. 31c). These data suggest that AGO2-Y393 phosphorylation decreases Dicer-AGO2 interaction, which in turn reduces miRNA precursor loading, suppresses the maturation of long-loop mHESM and decreases the loading of corresponding mature miRNAs onto RISC under hypoxia.

To demonstrate that the long-loop structure of precursor miRNAs indeed serves as one of the determinants that distinguish mHESM that are regulated by p-Y393-AGO2 from other miRNAs that are not, we mutated pre-miR-192-WT (long loop) into pre-miR-192-3M (short loop) and stably expressed them in HeLa Tet-Off-inducible AGO2 stable clones (Fig. 3f, top). Compared with AGO2-Y393F mutant, induction of AGO2-WT attenuated the maturation of pre-miR-192-WT but not pre-miR-192-3M, which virtually lost its processing efficacy without the long-loop structure (Fig. 3f and Supplementary Fig. 32). Conversely, AGO2-WT was able to suppress pre-miR-21-3M with a regenerated long-loop structure (Fig. 3g and Supplementary Fig. 33) that is not present in miR-21-WT (Fig. 3g and Supplementary Fig. 34). These results support a long-loop-dependent mechanism by which p-Y393-AGO2 confers regulation specificity on miRNA maturation.

The hypoxic tumour microenvironment promotes the metastatic phenotype by facilitating tumour cell survival through evasion of apoptosis¹⁸. Given that most mHESM suppressed by p-Y393-AGO2 are tumour-suppressor-like (Supplementary Figs 18 and 25b), we further investigated the pathophysiological role of AGO2 phosphorylation in response to hypoxia. Compared with vector control and AGO2-WT, a higher proportion of cells expressing AGO2-Y393F mutant underwent apoptosis following hypoxia exposure for three days (Supplementary Fig. 35), indicating that they were more susceptible to hypoxic stress. Knockdown of endogenous EGFR reduced cell survival and diminished the differences in apoptosis between AGO2-WT and AGO2-Y393F stable transfectants (Fig. 4a), suggesting that the phosphorylation of AGO2, rather than the mutation itself, is critical for cell survival under hypoxia. We did not observe any significant changes between AGO2-WT and AGO2-Y393F mutant in cell proliferation rate (Supplementary Fig. 36a, b) or anchorage-independent growth (Supplementary Fig. 36c, d). Notably, AGO2-WT but not AGO2-Y393F mutant significantly increased cell migration in response to hypoxia (Fig. 4b and Supplementary Figs 37 and 38). Treatment with Tyr-kinase inhibitor abrogated AGO2-WT-enhanced migration but failed to inhibit AGO2-Y393F mutant cells, indicating that AGO2-Y393 phosphorylation is important for EGFR-enhanced cell migration under hypoxia. Similar results were obtained from three-dimensional invasion assay with or without Tyr-kinase inhibitor treatment (Fig. 4c and Supplementary Fig. 39). These results demonstrate the functional importance of AGO2-Y393 phosphorylation in blocking cell apoptosis and enhancing cell invasiveness under hypoxia.

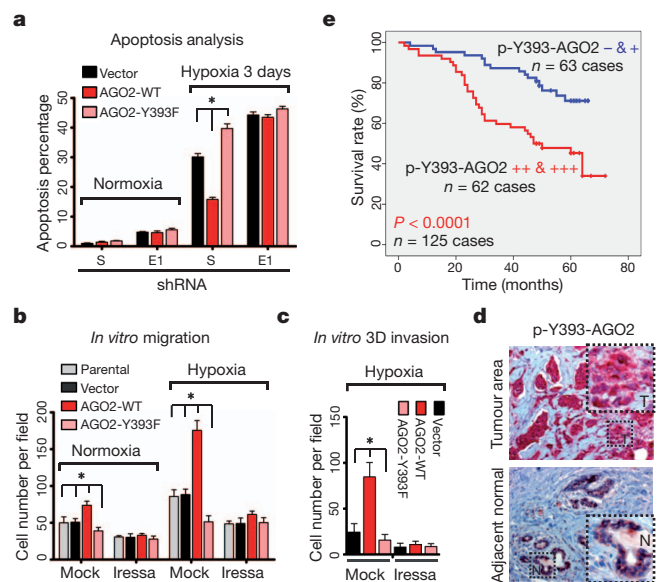


Figure 4 | p-Y393-AGO2 enhances cell survival and invasiveness under hypoxia and correlates with poorer overall survival in breast cancer patients.

a, Cell apoptosis analysed by FACS ($n = 3$). **b**, *In vitro* migration assay. Result was calculated on the basis of five randomly selected fields per membrane in triplicate ($n = 3$). **c**, *In vitro* three-dimensional invasion assay ($n = 5$). **d**, Representative images of immunohistochemical staining of p-Y393-AGO2 in human breast tumour and its adjacent normal tissue. **e**, Correlation between p-Y393-AGO2 and overall survival in breast cancer patients ($n = 125$). $P < 0.0001$, Kaplan-Meier survival analysis. All data represent mean \pm s.d.; * $P < 0.05$, *t*-test.

Finally, we used an orthotopic xenograft breast cancer model to establish the relationship between hypoxia, EGFR and p-Y393-AGO2, and showed that p-Y393-AGO2 along with EGFR is upregulated during tumour progression and specifically enriched in hypoxic tumour areas (Supplementary Fig. 40). To further examine the clinical relevance of AGO2-Y393 phosphorylation, we analysed the expression patterns of p-Y393-AGO2 and EGFR and the degree of hypoxia (indicated by HIF1 α and HIF2 α ; ref. 18) in primary breast tumours in consecutive sections collected from 128 human breast cancer patients. In adjacent normal breast tissues the expression of p-Y393-AGO2 was low, but in hypoxic breast tumours it was highly elevated (Fig. 4d). We observed significant positive correlations between p-Y393-AGO2, EGFR, HIF1 α and HIF2 α (Supplementary Fig. 41a and Supplementary Table 1) and further validated that AGO2-Y393 phosphorylation was enriched in hypoxic subareas of breast tumours with positive expression of EGFR (Supplementary Fig. 41b). Moreover, higher expression of p-Y393-AGO2 correlated significantly with poorer overall survival in breast cancer patients (Fig. 4e), supporting its clinical importance as a potential prognostic marker for breast cancer patient survival.

In this study, we identified a novel role for EGFR in miRNA maturation via AGO2-Y393 phosphorylation. Although these results suggest that EGFR is the Tyr kinase that suppresses miRNA maturation through AGO2-Y393 phosphorylation under hypoxia, there may be other Tyr kinases that can also contribute to phospho-AGO2-mediated miRNA processing. The work we present here opens a new direction to understanding further the regulation of miRNA machinery in response to stress signalling, which is likely to have important clinical implications.

METHODS SUMMARY

We used quantitative PCR with reverse transcription to measure the expression levels of precursor and mature miRNAs, as described previously²⁷. Customized next-generation RNA deep sequencing, including both small-RNA application and whole-transcriptome analysis, was performed according to the standard protocol (Applied Biosystems). The full methodology can be found in Supplementary Information.

Received 12 January 2012; accepted 19 March 2013.

Published online 1 May 2013.

1. Kim, V. N. MicroRNA biogenesis: coordinated cropping and dicing. *Nature Rev. Mol. Cell Biol.* **6**, 376–385 (2005).
2. van Kouwenhove, M., Kedde, M. & Agami, R. MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nature Rev. Cancer* **11**, 644–656 (2011).
3. Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435**, 834–838 (2005).
4. Leung, A. K. & Sharp, P. A. MicroRNA functions in stress responses. *Mol. Cell* **40**, 205–215 (2010).
5. Pouyssegur, J., Dayan, F. & Mazure, N. M. Hypoxia signalling in cancer and approaches to enforce tumour regression. *Nature* **441**, 437–443 (2006).
6. Gould, G. W. & Lippincott-Schwartz, J. New roles for endosomes: from vesicular carriers to multi-purpose platforms. *Nature Rev. Mol. Cell Biol.* **10**, 287–292 (2009).
7. Mosesson, Y., Mills, G. B. & Yarden, Y. Derailed endocytosis: an emerging feature of cancer. *Nature Rev. Cancer* **8**, 835–850 (2008).
8. Cikaluk, D. E. *et al.* GERp95, a membrane-associated protein that belongs to a family of proteins involved in stem cell differentiation. *Mol. Biol. Cell* **10**, 3357–3372 (1999).
9. Eulalio, A., Huntzinger, E. & Izaurralde, E. Getting to the root of miRNA-mediated gene silencing. *Cell* **132**, 9–14 (2008).
10. Diederichs, S. & Haber, D. A. Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. *Cell* **131**, 1097–1108 (2007).
11. Chendrimada, T. P. *et al.* TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**, 740–744 (2005).
12. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134 (2010).
13. Wang, Y. *et al.* Regulation of endocytosis via the oxygen-sensing pathway. *Nature Med.* **15**, 319–324 (2009).
14. Franovic, A. *et al.* Translational up-regulation of the EGFR by tumor hypoxia provides a nonmutational explanation for its overexpression in human cancer. *Proc. Natl Acad. Sci. USA* **104**, 13092–13097 (2007).
15. Reynolds, A. R., Tischer, C., Verveer, P. J., Rocks, O. & Bastiaens, P. I. EGFR activation coupled to inhibition of tyrosine phosphatases causes lateral signal propagation. *Nature Cell Biol.* **5**, 447–453 (2003).
16. Lee, O. H. *et al.* Genome-wide YFP fluorescence complementation screen identifies new regulators for telomere signaling in human cells. *Mol. Cell. Proteom.* **10**, M110.001628 (2010).
17. Jiang, X., Huang, F., Marusyk, A. & Sorkin, A. Grb2 regulates internalization of EGF receptors through clathrin-coated pits. *Mol. Biol. Cell* **14**, 858–870 (2003).
18. Bertout, J. A., Patel, S. A. & Simon, M. C. The impact of O₂ availability on human cancer. *Nature Rev. Cancer* **8**, 967–975 (2008).
19. Ventura, A. & Jacks, T. MicroRNAs and cancer: short RNAs go a long way. *Cell* **136**, 586–591 (2009).
20. Nicoloso, M. S., Spizzo, R., Shimizu, M., Rossi, S. & Calin, G. A. MicroRNAs – the micro steering wheel of tumour metastases. *Nature Rev. Cancer* **9**, 293–302 (2009).
21. Leung, A. K. & Sharp, P. A. MicroRNAs: a safeguard against turmoil? *Cell* **130**, 581–585 (2007).
22. Schirle, N. T. & MacRae, I. J. The crystal structure of human Argonaute2. *Science* **336**, 1037–1040 (2012).
23. Elkayam, E. *et al.* The structure of human Argonaute-2 in complex with miR-20a. *Cell* **150**, 100–110 (2012).
24. Tahbaz, N. *et al.* Characterization of the interactions between mammalian PAZ PIWI domain proteins and Dicer. *EMBO Rep.* **5**, 189–194 (2004).
25. Maniatakis, E. & Mourelatos, Z. A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev.* **19**, 2979–2990 (2005).
26. Tsutsumi, A., Kawamata, T., Izumi, N., Seitz, H. & Tomari, Y. Recognition of the pre-miRNA structure by *Drosophila* Dicer-1. *Nature Struct. Mol. Biol.* **18**, 1153–1158 (2011).
27. Suzuki, H. I. *et al.* Modulation of microRNA processing by p53. *Nature* **460**, 529–533 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank B. Pickering, D. Yu, and A.-B. Shyu for suggestions and technical assistance with northern blot analysis. This work was supported by the US National Institutes of Health (CA109311 and CA099031 to M.-C.H., and CCSG Core Grant CA16672), the US National Breast Cancer Foundation, The Center for Biological Pathway at the UT MD Anderson Cancer Center, S. G. Komen (SAC110016 to M.-C.H.), The Sister Institution Fund of China Medical University and Hospital and the UT MD Anderson Cancer Center, the Cancer Research Center of Excellence (DOH102-TD-C-111-005, Taiwan), a Private University grant (NSC99-2632-B-039-001-MY3, Taiwan), and the Program for Stem Cell and Regenerative Medicine Frontier Research (NSC101-2321-B-039-001, Taiwan).

Author Contributions J.S. and M.-C.H. designed and conceived the study; J.S. and M.-C.H. wrote the manuscript; J.L.H. contributed to the preparation of the manuscript. J.S., W.X., Y.B.K., L.H., S.-O.L., Y.D., Y. Wang, W.-C.C. and C.-H.C. did the experiments; Y. Wu provided human primary breast tumour samples; Y.C.L. provided the split-half-YFP-fused constructs; X.L. and C.-G.L. assisted in next-generation RNA deep sequencing; B.P.J. provided the pipeline analysis service for RNA sequencing data; and K.N. and D.-J.P. analysed the crystal structure of human AGO2.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.-C.H. (mhung@mdanderson.org).

Psl trails guide exploration and microcolony formation in *Pseudomonas aeruginosa* biofilms

Kun Zhao^{1*}, Boo Shan Tseng^{2*}, Bernard Beckerman³, Fan Jin⁴, Maxsim L. Gibiansky¹, Joe J. Harrison², Erik Luijten^{3,5}, Matthew R. Parsek² & Gerard C. L. Wong^{1,6,7}

Bacterial biofilms are surface-associated, multicellular, morphologically complex microbial communities^{1–7}. Biofilm-forming bacteria such as the opportunistic pathogen *Pseudomonas aeruginosa* are phenotypically distinct from their free-swimming, planktonic counterparts^{7–10}. Much work has focused on factors affecting surface adhesion, and it is known that *P. aeruginosa* secretes the Psl exopolysaccharide, which promotes surface attachment by acting as ‘molecular glue’^{11–15}. However, how individual surface-attached bacteria self-organize into microcolonies, the first step in communal biofilm organization, is not well understood. Here we identify a new role for Psl in early biofilm development using a massively parallel cell-tracking algorithm to extract the motility history of every cell on a newly colonized surface¹⁶. By combining this technique with fluorescent Psl staining and computer simulations, we show that *P. aeruginosa* deposits a trail of Psl as it moves on a surface, which influences the surface motility of subsequent cells that encounter these trails and thus generates positive feedback. Both experiments and simulations indicate that the web of secreted Psl controls the distribution of surface visit frequencies, which can be approximated by a power law. This Pareto-type¹⁷ behaviour indicates that the bacterial community self-organizes in a manner analogous to a capitalist economic system¹⁸, a ‘rich-get-richer’ mechanism of Psl accumulation that results in a small number of ‘elite’ cells becoming extremely enriched in communally produced Psl. Using engineered strains with inducible Psl production, we show that local Psl concentrations determine post-division cell fates and that high local Psl concentrations ultimately allow elite cells to serve as the founding population for initial microcolony development.

The extracellular matrix of *P. aeruginosa* biofilms is composed mostly of exopolysaccharides, proteins and DNA^{19–21}. Biofilms formed by non-mucoid isolates of *P. aeruginosa* primarily use two types of exopolysaccharide: Psl and Pel²². To initiate biofilm development, the wild-type strain PAO1 uses Psl to promote adhesion of cells to surfaces^{14,15}. Once associated with a surface, *P. aeruginosa* can remain fixed at the point of attachment or move. Formation of microcolonies (aggregates of ~50 cells or less) is the first step in the communal organization of a biofilm; however, little is known regarding the transition of individual cells to these discrete microcolonies^{3–5,23}.

Consistent with the role of Psl in surface adhesion^{7,24}, the average surface residence time of a PAO1 Δ pslD mutant strain (which cannot produce Psl) was $35 \pm 10\%$ shorter than that of wild type. Tracking algorithms allow us to isolate differences in spatial characteristics of cell–surface interactions, in addition to temporal characteristics such as residence time. We observed a fundamental difference in surface exploration patterns between the wild-type and Δ pslD strains (Fig. 1a–d and Supplementary Video 1) when we tracked the motility history of all cells in a $67 \mu\text{m} \times 67 \mu\text{m}$ field of view (>700,000 images of individual

cells, 11 h of data, time resolution of 3 s per frame). In Fig. 1a–d, black regions represent ‘untouched’ surface areas, whereas red regions represent areas visited by bacteria. For wild type, the surface coverage increased slowly to a maximum of $\sim 55 \pm 5\%$ in 5 h. In contrast, the Δ pslD mutant covered $79 \pm 10\%$ in 5 h. These observed differences are not due to differences in growth between strains (Supplementary Fig. 1). To confirm that changes in surface motility, rather than changes in the numbers of bacteria visiting the surface, are responsible for these observations, we also compared wild type and Δ pslD at the same total number of bacterial visits (that is, the sum of the number of bacteria in all frames, $N_s = \sum n_i$, where n_i is the number of bacteria in frame i ; Supplementary Fig. 2). Indeed, essentially the same trend was observed, with a surface coverage of $52 \pm 4\%$ for wild type at $N_s = 124,000$ and a surface coverage of $83 \pm 10\%$ for Δ pslD at the same N_s . These are averages from at least three replicates (Supplementary Methods).

Because Psl is important in both surface motility and adhesion, we characterized its spatial distribution with respect to bacterial trajectories. A TRITC-conjugated, Psl-specific lectin¹² was used to visualize Psl. We allowed wild-type cells to traverse a surface for 2 h before lectin staining and found that a trail of Psl was left on the surface (Supplementary Fig. 3). We verified that extracellular DNA²¹ was not found in the trails (by staining with propidium iodide (data not shown) and SYTO 9 (Supplementary Figs 4 and 5)).

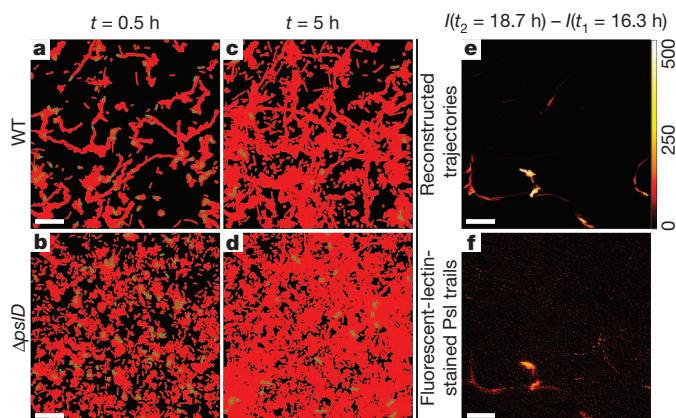


Figure 1 | Efficiency of surface coverage by bacterial trajectories and correlation with Psl trails. a–d, Cumulative surface coverage at 0.5 h (a, b) and 5 h (c, d). Red and black colours indicate traversed (that is, covered by bacterial trajectories) and fresh surface, respectively. Bacteria in the frame at the specified time points are shown in green. e, Reconstructed bacterial trajectories of wild type generated between 16.3 and 18.7 h after inoculation (colour scale indicates the time a given cell spent at each point). f, Psl trail left behind by bacteria in the same period, stained by fluorescently conjugated *Hippeastrum* hybrid (amaryllis) lectin. Scale bars, $10 \mu\text{m}$.

¹Department of Bioengineering, University of California, Los Angeles, California 90095, USA. ²Department of Microbiology, University of Washington, Seattle, Washington 98195, USA. ³Department of Materials Science and Engineering, Northwestern University, Evanston, Illinois 60208, USA. ⁴CAS Key Laboratory of Soft Matter Chemistry, University of Science and Technology of China, Hefei, Anhui 230026, China. ⁵Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, Illinois 60208, USA. ⁶Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, USA. ⁷California NanoSystems Institute, University of California, Los Angeles, California 90095, USA.

*These authors contributed equally to this work.

Cell-tracking algorithms and lectin staining were combined to investigate the spatiotemporal aspects of Psl deposition in a manner that accounts for the changing speeds (and, therefore, local residence times) of cells. Figure 1e shows bacterial trajectories between times t_1 (16.3 h after inoculation) and t_2 (18.7 h after inoculation). Figure 1f shows the lectin-stained Psl trails generated in this time period (Methods and Supplementary Fig. 6). Surface regions with the brightest fluorescence signals and, hence, the highest local surface concentrations of Psl corresponded to those with the longest bacterial visits. Multiple visits to a surface region by different cells and extended residence times of a specific cell on the same surface region both served to increase local Psl concentration.

We used cell-tracking algorithms to determine the bacterial visit frequency for each surface pixel for the entire community (Methods). A surface visit map (Fig. 2a) of all wild-type cells within the field of view (Fig. 2b, ~15.7 h after inoculation) shows that a considerable fraction of the surface was never visited. Of the surface areas traversed by bacteria, most were traversed once. In fact, the visit frequency distribution (histogram of the number of pixels with N bacterial visits) was measured to be a monotonically decreasing function of N . The precise form of the function is complex; however, for values of N ranging from a few visits to more than a hundred visits, the distribution is approximately

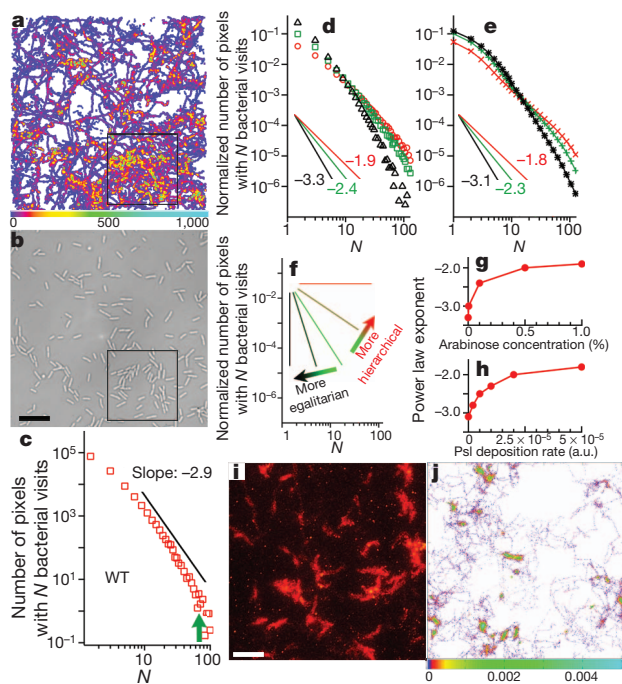


Figure 2 | Visit frequency distribution and its effect on bacterial movement. **a**, Visit frequency map of wild type for the first 15.7 h after inoculation, when microcolonies were just starting to form (example outlined by black square). **b**, Bright-field image for wild type at $t \approx 15.7$ h. **c**, Visit frequency distribution from **a**. The solid line shows a power-law decay with exponent -2.9 . The green arrow indicates where the curve begins to deviate from this power law. **d**, Visit frequency distributions for $\Delta P_{sl}/P_{BAD-psl}$ at arabinose concentrations of 0% (triangles), 0.1% (squares) and 1% (circles). **e**, Simulation results of visit frequency distributions at Psl deposition rates of 0 (asterisks), 10^{-5} (pluses), 5×10^{-5} (crosses) (arbitrary units; see Supplementary Methods). In **d** and **e**, each data set is normalized by the total number of visits (roughly the same as for **a**) and solid lines show power-law decay, labelled by exponent. **f**, Schematic graph showing that distributions with steep slopes are more egalitarian, whereas those with shallow slopes are more hierarchical. **g**, **h**, Fitted power-law exponents of visit frequency distributions from experiments at different arabinose concentrations (**g**) and simulations at different Psl deposition rates (**h**). **i**, A lectin-stained image showing hierarchical distribution of Psl ($\Delta P_{sl}/P_{BAD-psl}$ at 1% arabinose). **j**, Psl map from simulations (Psl deposition rate, 5×10^{-5} in dimensionless, arbitrary units (a.u.)). Scale bars, 10 μ m.

described by a power law with an exponent of -2.9 ± 0.1 (Fig. 2c), which serves as a simple metric for the distribution.

The importance of Psl in determining this global pattern of bacterial surface exploration can be seen in the visit histograms of the $\Delta P_{sl}/P_{BAD-psl}$ mutant strain (strain defined in Supplementary Methods), which uses an arabinose-inducible promoter to control Psl production²⁵. As the concentration of arabinose in the growth medium was changed from 0% to 1% (w/v), the visit histograms showed a systematic shift from a more uniform distribution of surface visits (that is, having a steep power-law decay) to a more hierarchical distribution (that is, having a broader range of visit frequencies) (Fig. 2d), as evidenced by the change in the effective power-law exponent from -3.3 ± 0.1 to -1.9 ± 0.1 , which correspond to low and high arabinose-induced Psl production, respectively. Because the total number of visits is the same for the different curves, a less steep power-law behaviour implies that with increasing Psl concentration the bacterial visits become concentrated in a smaller fraction of the surface. These observations lead us to propose that the higher the Psl concentration for a given surface region, the more likely it is that this region will be visited by more bacteria, causing further local accumulation of Psl and resulting in positive feedback.

Type IV pili (TFP) have a pivotal role in *P. aeruginosa* surface motility and the power-law exploratory mechanism. TFP can extend considerably from the cell body²⁶ and assist in ‘exploring’ surface areas beyond the cellular ‘footprint’. We propose that TFP associate with Psl-rich surface regions and pull the cell towards these regions with higher probability. Consistent with this, we show that a $\Delta pilA$ mutant strain, which eliminates TFP-driven twitching motility, results in a drastic reduction of surface exploration (Supplementary Fig. 7). Because cells can secrete Psl as well as associate with Psl, twitching motility effectively promotes positive feedback. Simple track following is just one possible manifestation of this mechanism: cells can approach a given track from any direction. Although such cells can move towards Psl, they do not necessarily reorient themselves tangentially to a given track.

Power-law relationships generally exist only over a limited range in nature and are often difficult to distinguish from other, quantitatively similar relationships. Therefore, rather than concentrating on the approximate power law in the visit frequency distribution and the range over which it is observed, we focused instead on obtaining a deeper quantitative understanding of the processes that generate its functional form. Thus, we performed computer simulations of Psl-guided motility of Psl-secreting bacteria using experimentally measured parameters to test stringently the quantitative interplay between progressive Psl secretion, surface motility and the idea of surface exploration guided by positive feedback.

Bacteria were modelled as non-overlapping line segments in a two-dimensional domain that, when unbiased by Psl, move according to a velocity distribution extracted from the experimental data for $\Delta P_{sl}/P_{BAD-psl}$ at 0% arabinose. In each step, a fixed amount of Psl was secreted and the bacterial displacement was biased by the local Psl distribution. The bacterial concentration and dimensions as well as the sampling rate were all chosen in accordance with the experimental parameters (Supplementary Methods). The simulations captured the complex distributions from experiments, in both the power-law regime and beyond (Figs 2e and 3c), including the dependence on Psl. The simulated visit frequency distributions exhibited a power-law behaviour that agrees quantitatively with the tracked microscopy data. This is a striking confirmation of the role of Psl-biased motion as the underlying mechanism, because unbiased motion would give rise to non-localized, random-walk-type behaviour. As the Psl production rate was increased for both experiments (Fig. 2d) and simulations (Fig. 2e), the exponents increased over the same numerical range, from -3.1 to -1.8 , confirming the shift to more hierarchical distributions in which the number of highly visited sites increased at the expense of many of the rarely visited sites (Fig. 2f–h). Likewise, both the lectin-stained

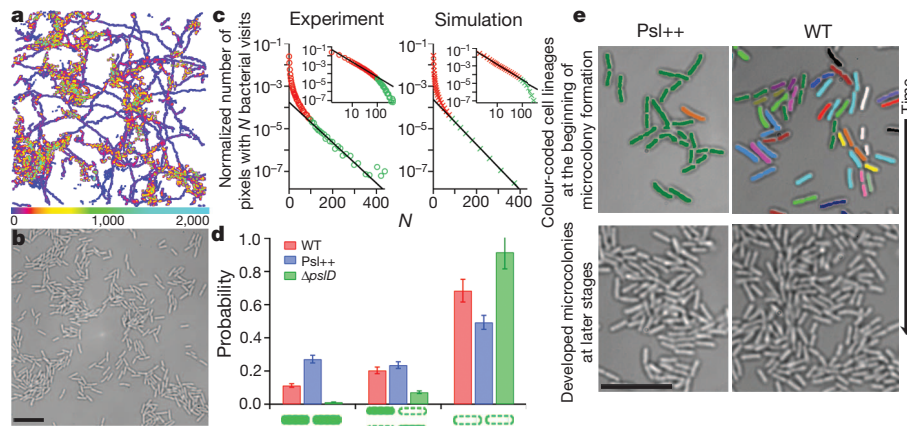


Figure 3 | Local Psl levels determine post-division cell fates. **a**, Visit frequency map of Psl+ for the first 14 h after inoculation. Microcolonies have already started to form. **b**, Bright-field image for Psl+ at $t \approx 14$ h. **c**, Visit frequency distributions of Psl+ from **a** for experiments (left) and simulations (right) agree. The solid line is an exponential fit to the second part of the data (green). Inset, power-law fit to the first part of the data (red). **d**, Probability of post-division cells' fates: stay (filled rod), leave (open dashed rod), wild type (red), Psl+ (blue), $\Delta pslD$ (green). Error bars are estimated from $1/\sqrt{N_{div}}$, where N_{div} is the total number of division events during the time period of interest ($N_{div} \geq 90$).

e, Wild-type and Psl+ microcolonies have drastically different compositions, as depicted by colour-coded cell division lineages at early stages of microcolony formation (top row). For Psl+ the microcolony is dominated by a single lineage, whereas the wild-type microcolony has 20 different lineages. The bottom row depicts more developed microcolonies at the same location 3.3 h later. Scale bars, 10 μ m.

Psl image observed in experiments (Fig. 2i) and the simulated Psl map (Fig. 2j) showed the hierarchical nature of the spatial distribution of Psl.

Power laws (such as Zipf's law^{17,18} or its continuous generalization, Pareto distributions) are known to describe self-organized systems like the wealth distributions in capitalist economies¹⁸. We found that bacteria are guided by synergistic rich-get-richer mechanisms. Both simulations and experiments indicate that the ability of *P. aeruginosa* to secrete Psl and its tendency to associate with Psl result in an exploratory strategy by which cells go where other cells go most often. (The range of exponents we observe here for bacteria is essentially the same as for income distributions²⁷.) This strategy results in a small number of cells being positioned at locations that are extremely enriched in communally produced Psl. This arrangement provides the conditions necessary for microcolony growth. Notably, these effects persist even when we completely turn off cell division in computer simulations, which indicates that the distribution of cells is a consequence of their exploratory strategy rather than of growth.

We examined the behaviour of bacterial communities beyond the power-law exploratory phase to show how growth can amplify differences in cell density on the surface caused by exploration. We found that high local Psl concentrations resulted in discrete microcolonies that exhibit localized exponential growth rather than delocalized surface exploration. For data on wild type (Fig. 2a–c), the power-law behaviour persisted only in the initial exploratory phase when there were no microcolonies. In fact, small deviations from the power law can be observed in the lower right portion of Fig. 2c (green arrow), which corresponds to highly visited areas with the highest local Psl concentrations. To examine community behaviour near high Psl concentrations, we examined the $\Delta P_{psl}/P_{BAD-psl}$ mutant that overproduces Psl in the presence of 1% arabinose (Psl+). The bacterial visit frequency map for the Psl+ cells was significantly different from that of wild-type cells (Fig. 3a). Compared with Fig. 2a, there are far fewer tracks leading into and out of the cellular aggregates in Fig. 3a, implying that few cells joined or departed from these aggregates. The regions of highest numbers of bacterial visits in the map fully correlate with the existence of microcolonies of closely packed bacteria in the bright-field image (Fig. 3b). The cell densities in the highly visited areas were more than three times higher than the average cell density on the surface. Moreover, the Psl+ cells formed microcolonies earlier than did the wild-type cells. Notably, owing to microcolony formation, a portion of the visit frequency distribution is

described by an exponential rather than a power law in both experiments and simulations (Fig. 3c).

We used cell tracking to show that the exponential growth which enables localized microcolony formation is ultimately rooted in how local Psl concentrations determine post-division cell fates. We distinguished three possibilities: the post-division cells both stay near the division event, both leave (see definitions of 'stay' and 'leave' in Supplementary Methods) or one cell stays and the other leaves. Figure 3d shows the probabilities of these cell fates observed for wild type, Psl+ and $\Delta pslD$. The probability of both cells staying for the Psl+ mutant was approximately three times as high as for wild type. In contrast, the probability of both cells staying for the $\Delta pslD$ mutant strain was essentially zero. In fact, the fraction of total population growth due to cell division (as opposed to motile cells joining the microcolony, cells attaching from solution and so on) was $\sim 85\%$ for the Psl+ mutant, which directly leads to an exponential visit distribution (Supplementary Fig. 8). We traced the division lineage of cells in developing microcolonies (Fig. 3e). The Psl+ microcolony was derived mostly from a single lineage, consistent with our proposal that such cells primarily divide and remain in the same area owing to the high local Psl concentrations. In contrast, the wild-type microcolony developed from many cell lineages as a result of the initial power-law exploratory phase. (More examples can be found in Supplementary Figs 9 and 10.) These data strongly support our rich-get-richer model for microcolony formation.

The results presented here quantitatively show the key steps important for the transition from surface attachment and surface exploration to microcolony formation: the evolving distribution of Psl on a solid surface provides a mechanism for bacteria to self-organize socially, resulting in a rich-get-richer power-law distribution of visit frequencies and a hierarchical distribution of Psl concentrations on the surface. This in turn results in a small number of aggregated cells becoming highly enriched in Psl, which enables them to serve as the founding population for localized exponential growth and the formation of microcolonies that ultimately evolve into the mushroom-like 'stalks-and-caps' morphology of mature biofilms^{4–6}.

METHODS SUMMARY

Wild-type *P. aeruginosa* strain PAO1²⁸ and its isogenic mutants $\Delta pslD$, $\Delta P_{psl}/P_{BAD-psl}$ ¹⁴ and $\Delta pilA$ were used in this study. The growth rates of the strains are nearly identical (to within measurement error; Supplementary Fig. 1). For $\Delta P_{psl}/P_{BAD-psl}$, arabinose was used to control Psl production¹⁴. FAB medium²⁹

with 0.6 mM glutamate was used for flow cell experiments, which were conducted at 30 °C with a flow rate of 3 ml h⁻¹. Flow cells were assembled as previously described³⁰. The flow was stopped for bacterial inoculation and 20 min of incubation, and then resumed with the initiation of image recording.

Images were made with an Olympus microscope. Bright-field images were recorded every 3 s. Each data set, which had 18,000–48,000 frames, contained up to 1,000,000 bacteria images. Psl was stained with 100 µg ml⁻¹ TRITC-labelled *Hippeastrum* hybrid lectin from amaryllis¹². The image size was 67 µm × 67 µm (1,024 × 1,024 pixels).

In simulations, line segments (representing bacteria) moved within a square surface window with periodic boundary conditions. Each bacterium moved with fixed step size at a variable angle deviating from its cell body axis according to distributions drawn from experiments, and deposited Psl at its centre at a tunable rate. This Psl in turn modified the bacterial motion through Psl-concentration-based biasing. In addition, random reorientations happened at a prescribed rate in accordance with the bacterial mean square displacement from experiments, influenced by Psl. The total simulation time was chosen such that the final visit count was the same as in experiments.

Full Methods and any associated references are available in the online version of the paper.

Received 19 September 2012; accepted 8 April 2013.

Published online 8 May 2013.

- Costerton, J. W., Stewart, P. S. & Greenberg, E. P. Bacterial biofilms: a common cause of persistent infections. *Science* **284**, 1318–1322 (1999).
- O'Toole, G. A. & Kolter, R. Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. *Mol. Microbiol.* **30**, 295–304 (1998).
- O'Toole, G., Kaplan, H. B. & Kolter, R. Biofilm formation as microbial development. *Annu. Rev. Microbiol.* **54**, 49–79 (2000).
- Stoodley, P., Sauer, K., Davies, D. G. & Costerton, J. W. Biofilms as complex differentiated communities. *Annu. Rev. Microbiol.* **56**, 187–209 (2002).
- Klausen, M., Aaes-Jørgensen, A., Molin, S. & Tolker-Nielsen, T. Involvement of bacterial migration in the development of complex multicellular structures in *Pseudomonas aeruginosa* biofilms. *Mol. Microbiol.* **50**, 61–68 (2003).
- Davies, D. G. *et al.* The involvement of cell-to-cell signals in the development of a bacterial biofilm. *Science* **280**, 295–298 (1998).
- Mann, E. E. & Wozniak, D. J. *Pseudomonas* biofilm matrix composition and niche biology. *FEMS Microbiol. Rev.* **36**, 893–916 (2012).
- Lyczak, J. B., Cannon, C. L. & Pier, G. B. Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist. *Microbes Infect.* **2**, 1051–1060 (2000).
- Tattersson, L. E., Poschet, J. F., Firoved, A., Skidmore, J. & Deretic, V. CFTR and pseudomonas infections in cystic fibrosis. *Front. Biosci.* **6**, d890–897 (2001).
- Singh, P. K. *et al.* Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms. *Nature* **407**, 762–764 (2000).
- Friedman, L. & Kolter, R. Two genetic loci produce distinct carbohydrate-rich structural components of the *Pseudomonas aeruginosa* biofilm matrix. *J. Bacteriol.* **186**, 4457–4465 (2004).
- Ma, L. Y., Lu, H. P., Sprinkle, A., Parsek, M. R. & Wozniak, D. J. *Pseudomonas aeruginosa* Psl is a galactose- and mannose-rich exopolysaccharide. *J. Bacteriol.* **189**, 8353–8356 (2007).
- Byrd, M. S. *et al.* Genetic and biochemical analyses of the *Pseudomonas aeruginosa* Psl exopolysaccharide reveal overlapping roles for polysaccharide synthesis enzymes in Psl and LPS production. *Mol. Microbiol.* **73**, 622–638 (2009).
- Ma, L., Jackson, K. D., Landry, R. M., Parsek, M. R. & Wozniak, D. J. Analysis of *Pseudomonas aeruginosa* conditional psl variants reveals roles for the psl polysaccharide in adhesion and maintaining biofilm structure postattachment. *J. Bacteriol.* **188**, 8213–8221 (2006).
- Byrd, M. S., Pang, B., Mishra, M., Swords, W. E. & Wozniak, D. J. The *Pseudomonas aeruginosa* exopolysaccharide Psl facilitates surface adherence and NF-κB activation in A549 cells. *mBio* **1**, e00140–10 (2010).
- Gibiansky, M. L. *et al.* Bacteria use type IV pili to walk upright and detach from surfaces. *Science* **330**, 197 (2010).
- Newman, M. E. J. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **46**, 323–351 (2005).
- Gabaix, X. Power laws in economics and finance. *Annu. Rev. Econ.* **1**, 255–294 (2009).
- Eagon, R. G. Composition of an extracellular slime produced by *Pseudomonas aeruginosa*. *Can. J. Microbiol.* **8**, 585–586 (1962).
- Sutherland, I. W. The biofilm matrix: an immobilized but dynamic microbial environment. *Trends Microbiol.* **9**, 222–227 (2001).
- Whitchurch, C. B., Tolker-Nielsen, T., Ragas, P. C. & Mattick, J. S. Extracellular DNA required for bacterial biofilm formation. *Science* **295**, 1487 (2002).
- Colvin, K. M. *et al.* The Pel and Psl polysaccharides provide *Pseudomonas aeruginosa* structural redundancy within the biofilm matrix. *Environ. Microbiol.* **14**, 1913–1928 (2012).
- Monroe, D. Looking for chinks in the armor of bacterial biofilms. *PLoS Biol.* **5**, e307 (2007).
- Flemming, H. C., Neu, T. R. & Wozniak, D. J. The EPS matrix: the “house of biofilm cells”. *J. Bacteriol.* **189**, 7945–7947 (2007).
- Newman, J. R. & Fuqua, C. Broad-host-range expression vectors that carry the l-arabinose-inducible *Escherichia coli* araBAD promoter and the araC regulator. *Gene* **227**, 197–203 (1999).
- Skerker, J. M. & Berg, H. C. Direct observation of extension and retraction of type IV pili. *Proc. Natl Acad. Sci. USA* **98**, 6901–6904 (2001).
- Solomon, S. & Richmond, P. Power laws of wealth, market order volumes and market returns. *Physica A* **299**, 188–197 (2001).
- Holloway, B. W. Genetic recombination in *Pseudomonas aeruginosa*. *J. Gen. Microbiol.* **13**, 572–581 (1955).
- Heydorn, A. *et al.* Experimental reproducibility in flow-chamber biofilms. *Microbiology* **146**, 2409–2415 (2000).
- Sternberg, C. & Tolker-Nielsen, T. Growing and analyzing biofilms in flow cells. *Curr. Protocols Microbiol.* **21**, 1B.2.1–1B.2.17 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements K.Z., B.S.T., M.R.P. and G.C.L.W. are supported by the US National Institutes of Health (NIH 1R01HL087920). K.Z. and G.C.L.W. also acknowledge support from the US National Science Foundation (NSF DMR1106106) and a UCLA Transdisciplinary Seed Grant. B.S.T., J.J.H. and M.R.P. also acknowledge support from the NIH (R01AI077628, R01AI081983, R56AI061396) and NSF (MCB0822405). B.S.T. is supported by the Cystic Fibrosis Foundation Postdoctoral Fellowship (TSENG11F0). J.J.H. was supported by a postdoctoral fellowship from the Natural Sciences and Engineering Research Council of Canada. B.B. and E.L. acknowledge support from the NSF under DMR-1006430 (E.L.) and DGE-0824162 (B.B.). The authors would like to thank J. Copic for discussions and R. J. Siehnell for technical assistance. We dedicate this paper to the memory of M. Shannon.

Author Contributions G.C.L.W., M.R.P. and K.Z. conceived the project. K.Z., B.S.T., M.R.P. and G.C.L.W. designed studies. K.Z. and B.S.T. performed experimental measurements. K.Z. and G.C.L.W. performed image analysis. F.J. helped in performing image analysis. M.L.G. helped in collecting experimental data. B.S.T., J.J.H. and M.R.P. constructed strains. B.B. and E.L. designed the model and performed computer simulations. K.Z., B.S.T., B.B., E.L., M.R.P. and G.C.L.W. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.C.L.W. (gclwong@seas.ucla.edu), M.R.P. (parsem@u.washington.edu) or E.L. (luijten@northwestern.edu).

METHODS

Strains and growth conditions. *Pseudomonas aeruginosa* strain PAO1 wild type²⁸ and its isogenic mutants $\Delta pslD$, the *psl*-inducible strain $\Delta P_{psl}/P_{BAD-psl}$ and $\Delta pilA$ were used in this study. The growth rates of the strains were nearly identical (to within measurement error; Supplementary Fig. 1). For strain $\Delta P_{psl}/P_{BAD-psl}$, different amounts of arabinose (Sigma-Aldrich; 1%, 0.1% and 0.005% were used in this work) were added into the medium to control the production of Psl¹⁴.

FAB medium²⁹ with 0.6 mM glutamate (Sigma-Aldrich) as the carbon source was used for flow cell experiments. An inoculum was prepared by growing strains in test tubes containing FAB with 30 mM glutamate, shaking at 220 r.p.m. and 37 °C to $D_{600nm} \sim 0.5$. Cultures were then diluted to $D_{600nm} \sim 0.01$ in FAB medium with 0.6 mM glutamate, which was used for injection into the flow chamber.

Flow cell assembly, sterilization and washing of the system. The flow cell was purchased from the Department of Systems Biology, Technical University of Denmark and assembled as previously described³⁰. The assembled flow cell was connected to a syringe through a 0.22- μm filter (Fisher Scientific) using silicon tubing (Dow Corning). Then the whole system was sterilized overnight with 3% H₂O₂ (Fisher Scientific) at 3 ml h⁻¹ using a syringe pump (KD Scientific). After the sterilization, autoclaved, deionized water was used to wash the whole system overnight. The system was washed again using FAB medium with 0.6 mM glutamate before starting the inoculation of the bacteria into the flow cell. The flow was stopped for bacterial inoculation and 20 min of incubation, and then resumed with the initiation of image recording. Different flow rates have recently been reported to have an effect on bacterial surface behaviour³¹, and it will be interesting to using large-scale tracking methods in this context. In this work, the flow cell experiments were conducted at 30 °C with a flow rate of 3 ml h⁻¹.

Data acquisition. The images were made using an EMCCD camera (Andor iXon) using IQ software (Andor) on an Olympus IX81 microscope equipped with Zero Drift autofocus system. The bright-field images were recorded every 3 s for a total recording time of about 20 h (varying by strain). Each data set, which had 18,000–48,000 frames, contained up to 1,000,000 bacteria images. The image size is 67 $\mu m \times 67 \mu m$ (1,024 \times 1,024 pixels).

Lectin staining. The Psl trails left by bacteria were stained with TRITC-labelled *Hippeastrum* hybrid (amaryllis) (HHA) lectin (EY laboratories)¹². Flow was suspended and 0.3 ml of 100 $\mu g ml^{-1}$ TRITC-HHA in FAB medium with 0.6 mM glutamate was injected upstream of the inlet flow. During the injection, the flow

chamber was monitored through the camera to make sure there was no shift or disturbance of the stage due to injection. The flow chamber was stained for 15 min without flow and then washed for 15 min with flow at a flow rate of 3 ml h⁻¹ in the dark before imaging.

Image analysis, surface coverage map and visit frequency map. Images were analysed as previously described¹⁶. In Fig. 1a–e, we used the area of each moving cell to generate the surface coverage from the bacterial trajectories. The visit frequency distributions are generated at full data resolution; for the visit maps, we used the centre of the cell to mark the trail, and spread each of these marks over a square patch eight pixels ($\sim 0.5 \mu m$) wide, in accordance with bacterium width. The distribution of Psl on the PAO1 surface is expected to be more complex³², but this will amount to a more complex point spread function and is not expected to alter the results qualitatively.

Simulation of Psl-guided motility of Psl-secreting bacteria. The computer simulations used a two-dimensional model in which an exponentially growing number of line segments of unit length moved within a square, periodic domain of side length 35, approximately corresponding to the experimental viewing window. Each bacterium was assigned a fixed step size and stepped with a variable angle. Both the step size and the stepping angle were drawn from the experimentally determined motion distribution for the $\Delta P_{psl}/P_{BAD-psl}$ strain at an arabinose concentration of 0% (that is, in the absence of Psl). Starting from a random non-overlapping configuration, all bacteria propagated in discrete time steps ($\Delta t = 3$ s), either moving at their constant, assigned rate, keeping their director fixed, or randomly reorienting. Reorientations occurred at a frequency 3.7%, determined by the mean squared displacement of the bacteria in the experiment. The effect of arabinose was represented by letting each bacterium deposit Psl at its centre at a tunable rate. The surface accumulation of Psl was recorded in elementary squares of side length 1/29 (corresponding to the experimental pixel size). The motion was in turn biased according to the presence of Psl within the vicinity of the leading pole of the bacterium. The total simulation time was chosen such that the final visit count was the same as in experiments. Over this window, we sampled individual bacterial trajectories as well as visit frequencies of all pixels (Supplementary Methods).

31. Lecuyer, S. *et al.* Shear stress increases the residence time of adhesion of *Pseudomonas aeruginosa*. *Biophys. J.* **100**, 341–350 (2011).
32. Ma, L. *et al.* Assembly and development of the *Pseudomonas aeruginosa* biofilm matrix. *PLoS Pathog.* **5**, e1000354 (2009).

Receptor binding by a ferret-transmissible H5 avian influenza virus

Xiaoli Xiong^{1*}, Peter J. Coombs^{1*}, Stephen R. Martin^{1*}, Junfeng Liu², Haixia Xiao^{1†}, John W. McCauley¹, Kathrin Locher³, Philip A. Walker¹, Patrick J. Collins¹, Yoshihiro Kawaoka⁴, John J. Skehel¹ & Steven J. Gamblin¹

Cell-surface-receptor binding by influenza viruses is a key determinant of their transmissibility, both from avian and animal species to humans as well as from human to human. Highly pathogenic avian H5N1 viruses that are a threat to public health have been observed to acquire affinity for human receptors, and transmissible-mutant-selection experiments have identified a virus that is transmissible in ferrets^{1–3}, the generally accepted experimental model for influenza in humans. Here, our quantitative biophysical measurements of the receptor-binding properties of haemagglutinin (HA) from the transmissible mutant indicate a small increase in affinity for human receptor and a marked decrease in affinity for avian receptor. From analysis of virus and HA binding data we have derived an algorithm that predicts virus avidity from the affinity of individual HA–receptor interactions. It reveals that the transmissible-mutant virus has a 200-fold preference for binding human over avian receptors. The crystal structure of the transmissible-mutant HA in complex with receptor analogues shows that it has acquired the ability to bind human receptor in the same folded-back conformation as seen for HA from the 1918, 1957 (ref. 4), 1968 (ref. 5) and 2009 (ref. 6) pandemic viruses. This binding mode is substantially different from that by which non-transmissible wild-type H5 virus HA binds human receptor. The structure of the complex also explains how the change in preference from avian to human receptors arises from the Gln226Leu substitution, which facilitates binding to human receptor but restricts binding to avian receptor. Both features probably contribute to the acquisition of transmissibility by this mutant virus.

The importance of HA, which mediates virus binding to cell-surface sialic acid moieties in the multi-gene-dependent process of H5N1 influenza virus transmission⁷, has been examined in several recent studies of ferrets infected with receptor-binding transmissible mutants^{1–3}. We have studied the previously described transmissible mutant¹ that acquired the ability to transmit in respiratory droplets, without contact, from ferret to ferret. The transmissible-mutant HA contained three amino acid substitutions in or near the receptor-binding site and a fourth, about 70 Å from it, nearer to the virus membrane (Supplementary Figs 1 and 2). One of the mutations, Gln226Leu, is shared by HAs of the other droplet-transmissible H5N1 viruses recently described and was key to the acquisition of human transmissibility by the 1957 (H2) and 1968 (H3) pandemic viruses^{8,9}. Such transmissibility from avian to human requires a change in binding preference from sialic acid in α 2,3-linkage to galactose on carbohydrate side chains (characteristic of the virus receptors in avian enteric tracts) to sialic acid in α 2,6-linkage (characteristic of human trachea airway epithelia)^{10,11}.

We have used two procedures to quantify the receptor-binding affinity and specificity of H5 viruses: microscale thermophoresis (MST) using recombinant HA trimers, and surface biolayer interferometry (BLI) with purified viruses. We have established a simple algorithm that enables virus binding to surfaces to be predicted from the dissociation

constant ($K_{d(\text{receptor})}$) of a single receptor interaction (Supplementary Fig. 2). Several key findings emerge. First, virus avidity for a given receptor is proportional to $K_{d(\text{receptor})}$ to the power of the multiplicity coefficient, the factor that accounts for the contributions of multiple HA–receptor interactions that occur between a virus and a target membrane. This relationship explains how typically millimolar $K_{d(\text{receptor})}$ values for HA affinity for its receptor can propagate to the femtomolar avidities observed for virus binding. Crucially, our data show that there is an effective upper value for the multiplicity coefficient of about 5.5, which derives from the geometry of virus particles¹² and the thermodynamics of the interactions. This limitation on virus valency explains how relatively modest decreases in $K_{d(\text{receptor})}$, perhaps 15-fold, can lead to an almost complete loss of virus binding and accounts for the biological observations of receptor-binding specificity (Supplementary Fig. 2).

Our MST data (Fig. 1a) show that the affinity of transmissible-mutant HA for human receptor is slightly increased in comparison to HA from the wild-type A/Vietnam/1194/2004 (H5N1) (VN1194) strain ($K_d = 12$ mM versus 17 mM), whereas the affinity for avian receptor is greatly decreased ($K_d = 32$ mM versus 1.1 mM). The data also show that the transmissible-mutant HA binds about fivefold weaker to human receptor, and about tenfold weaker to avian receptor, than HA from the well-characterized 1968 pandemic H3 virus A/Aichi/2/68 (H3N2), vaccine strain X-31. The binding constants for the latter, determined here by MST, for both receptor types are in good agreement with those determined previously by NMR¹³.

Our interferometry data for wild-type VN1194 H5 virus binding to human and avian receptor analogues attached to the biosensor are given in Fig. 1b, together with the data for X-31. These data show that the wild-type VN1194 H5 virus has very strong avidity towards avian receptor, as is found for nearly all avian viruses^{4,8,14}, and, unusually for an avian virus, almost none towards human receptor. The discrimination is such that the preference ratio (human/avian) for wild-type VN1194 H5 can only be estimated to be poorer than 5×10^{-7} . By contrast, X-31 has a preference ratio of about 10, which is similar to other pandemic viruses. To obtain information for the transmissible-mutant virus we used MST data for its HA and our empirical algorithm that relates HA affinity to virus avidity.

In Fig. 1c we present the binding curves predicted for wild-type VN1194 and transmissible-mutant viruses from the MST data using the algorithm. Importantly, there is close agreement between the experimental and predicted virus binding curves for wild-type VN1194 H5, which gives confidence in the predicted behaviour of the transmissible-mutant virus. Figure 1c shows that the transmissible-mutant virus has a preference ratio of about 200 in favour of human receptor binding, which is achieved by modest avidity for human receptor being offset by essentially undetectable binding to avian receptor. Qualitatively similar results to these have been reported previously¹. For comparison, transmissible-mutant virus binding to human receptor is $\sim 10^4$ times

¹MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK. ²Ministry of Agriculture, Key Laboratory of Plant Pathology, China Agricultural University, Yuanmingyuanxilu, 2, Beijing 100193, China. ³Novartis Institutes for BioMedical Research, Klybeckstrasse 141, CH-4057 Basel, Switzerland. ⁴Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, Wisconsin 53711, USA. [†]Present address: Laboratory of Protein Engineering and Vaccines, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China.

*These authors contributed equally to this work.

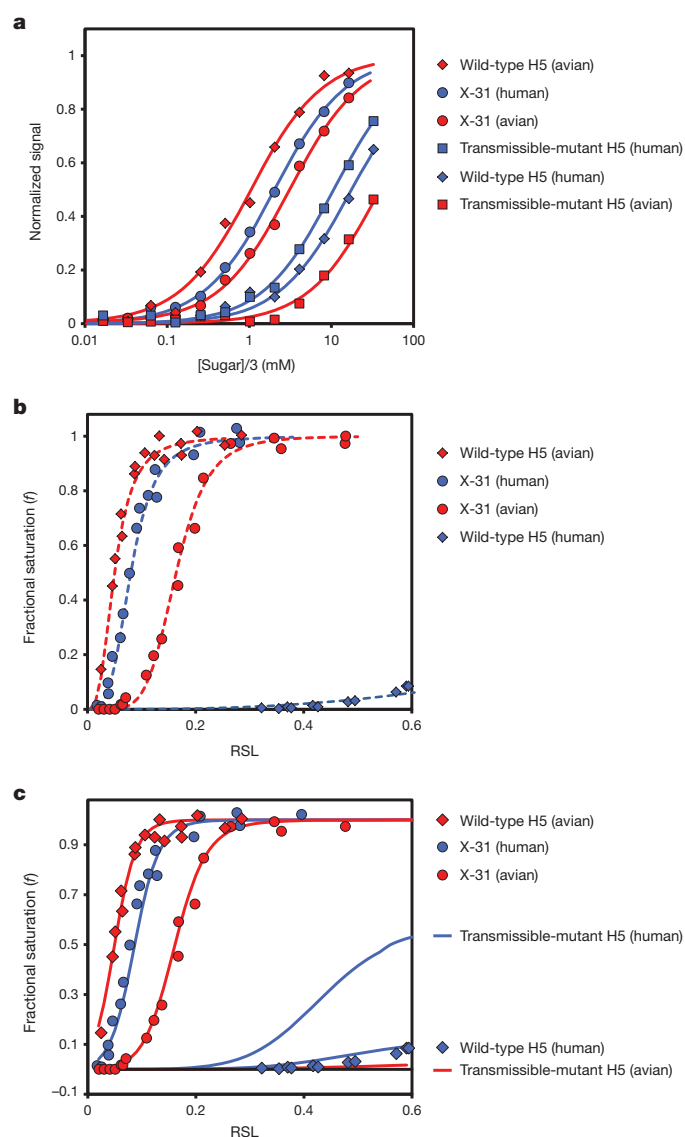


Figure 1 | HA and virus binding to human and avian receptors. **a**, MST data for the binding of α 2,3-linked sialyl lactosamine (α 2,3-SLN) and α 2,6-linked sialyl lactosamine (α 2,6-SLN) to HAs. Data are plotted as normalized signal change as a function of receptor sugar concentration divided by 3. For a full explanation of the normalization of the sugar concentration see Supplementary Information. **b**, BLI data for the binding of viruses to α 2,3-SLN and α 2,6-SLN. 30-kDa polymers containing 20% mol sugar and 5% mol biotin linked to a polyacrylamide backbone were immobilized to different levels on streptavidin-coated biosensors. Data are plotted as fractional saturation of the sensor surface as a function of relative sugar loading (RSL) for a fixed virus concentration of 100 pM. **c**, Comparison of experimental and simulated BLI data. The symbols represent the BLI data for the wild-type H5 virus. The solid lines are binding curves simulated using the HA monomer dissociation constants derived from MST experiments and the multiplicity coefficient (see Supplementary Information).

weaker than X-31, whereas its binding to avian receptor is 10^5 – 10^6 -times weaker than X-31.

We estimate that the highest receptor loadings achieved on the BLI biosensors are of the same order as accessible sialylated glyco-conjugates on target cell surfaces (Supplementary Fig. 2). Viruses that barely bind to the biosensor at this concentration are unlikely to be able to bind effectively to cells. It seems clear, therefore, that although the transmissible-mutant virus binds human receptor more weakly than pandemic viruses, about half of the virus will be bound at the receptor densities found on cell surfaces, a conclusion consistent with its observed transmissibility. On the other hand the transmissible-mutant

virus would not bind effectively to cells with predominantly avian receptors. As a consequence the specificity properties of the transmissible-mutant virus would favour infection of cells in the human upper respiratory tract that have been shown to display α 2,6-linked sialic acids. At the same time, the loss of avian receptor binding would also prevent the sequestration of the virus by airway mucins, which are rich sources of α 2,3-linked sialic acid¹⁵. Avoidance of mucins is probably important for the ferret transmissibility of the transmissible-mutant virus given that its human-receptor-binding properties are modest compared to pandemic viruses.

To understand the molecular basis of the altered affinity and specificity of the transmissible-mutant HA we determined its crystal structure, and that of wild-type VN1194 (clade 1) and A/turkey/Turkey/1/2005 (H5N1) (tyTy) (clade 2) H5 HAs in complex with human receptor analogues (Fig. 2, Supplementary Table 1 and Supplementary Fig. 1). The most notable feature of these structures is the different orientation of galactose at position 2 (Gal-2) and *N*-acetylglucosamine at position 3 (GlcNAc-3) of the receptor in the transmissible-mutant HA complex compared with their arrangement in both wild-type HA complexes (Fig. 2). In the transmissible-mutant complex, Gal-2 adopts a *cis* configuration about the glycosidic bond with sialic acid, and GlcNAc-3 exits the receptor-binding site towards the 190-helix (residues 187–197) (Fig. 2a, b). This orientation of Gal-2 and GlcNAc-3 is very similar to that in complexes formed between human receptor and HAs from the 1918 (H1), 1957 (H2) and 1968 (H3) pandemic viruses^{4,5} (Fig. 3a). However, it is in marked contrast to that seen in the wild-type H5 HA structures (Fig. 2e, g). Although only the first two sugars are ordered in the VN1194 complex, compared with three in the tyTy complex, sialic acid at position 1 of the receptor (Sia-1) and Gal-2 adopt the same conformation in both. In these cases (and all other mutant H5 HAs we have examined (X.X. *et al.*, unpublished observations)), Gal-2 also adopts a *cis* configuration about the glycosidic bond with sialic acid but it is rotated by approximately 90° about its C6–C5 bond such that GlcNAc-3 exits from the side of the receptor-binding site, over the 130-loop (residues 132–138). This mode of human-receptor binding has not been reported for any other avian or human HA. It seems to be enabled by the formation of a hydrogen bond between the 3' hydroxyl of Gal-2 with the main-chain carbonyl at residue 225 (Fig. 3b). This interaction is facilitated by the low position in the site adopted by Gln 226, which is stabilized by a hydrogen bond between its side chain and the hydroxyl of Ser 137.

The most obvious explanation for the difference in Gal-2 orientation in the complex formed by the transmissible-mutant HA with the human receptor is the presence at the base of the receptor-binding site of the hydrophobic leucine residue at position 226, in place of the polar glutamine residue found in wild-type H5 HAs. The structural effect of this Gln226Leu mutation in the transmissible mutant appears similar to that caused by the same substitution associated with the conversion of avian H2 and H3 to human viruses. In all three cases the introduction of the hydrophobic residue increases the separation between the 220- and 130-loops by about 1 Å, and residues 224 and 225 on the 220-loop (residues 220–228) sit about 1.4 Å higher in the site as a consequence of human-receptor binding (Supplementary Fig. 1).

To understand the loss of avian-receptor binding by transmissible-mutant HA we determined its structure, and that of the wild-type H5 HAs, in complex with an avian receptor analogue (Fig. 2c, f, h and Supplementary Fig. 1k–o). The most notable feature of these complexes is that, whereas both wild-type HAs bind avian receptor in the typical fashion of all avian HAs reported to date^{4,16,17} (Fig. 2f, h), the transmissible-mutant HA binds in a substantially different way with Gal-2 adopting a *cis*, rather than a *trans*, conformation about the glycosidic bond with sialic acid (Fig. 2c). As a result, Gal-2 appears face-on in the transmissible-mutant complex in Fig. 2 whereas it is edge-on in the wild-type H5 HA complexes. The altered orientation of Gal-2 and of GlcNAc-3 also results in a quite different trajectory for the avian receptor in the transmissible-mutant HA complex compared

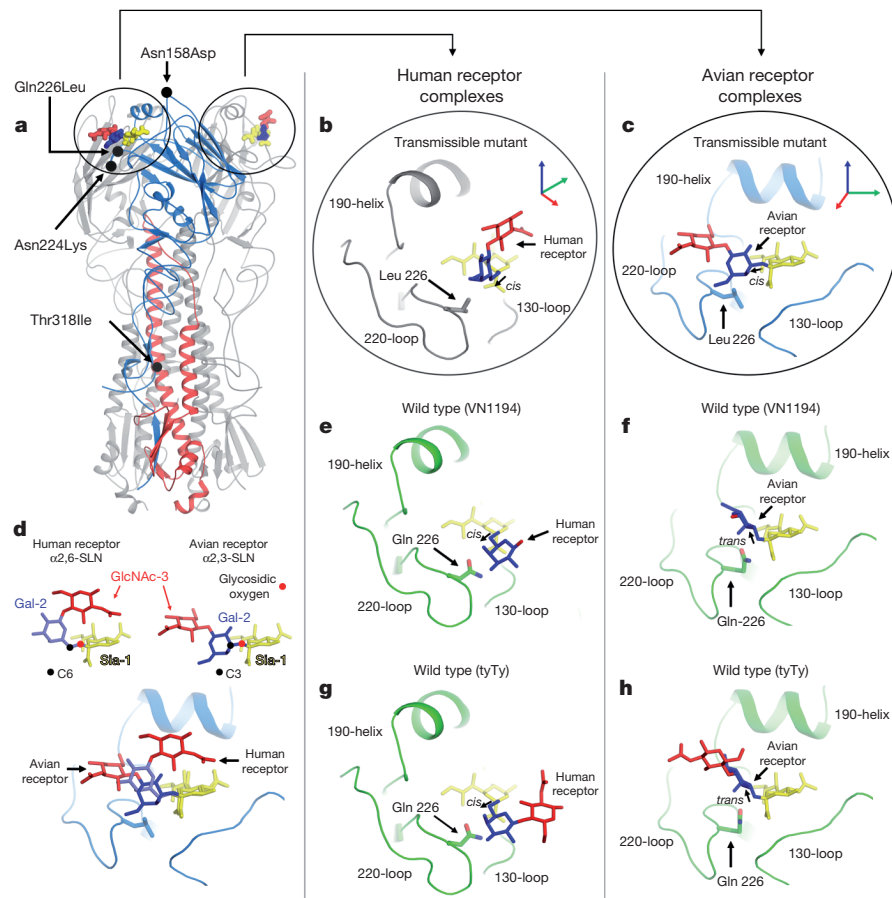


Figure 2 | Crystal structures of receptor complexes of wild-type H5 and transmissible-mutant HAs. **a**, Ribbons representation of an HA trimer of transmissible mutant with one monomer coloured in blue (HA-1) and red (HA-2), and the other two monomers shown in grey. For two of the monomers the structures of the bound receptor analogues are encircled and magnified in **b** (human receptor) and **c** (avian receptor). The locations of the four mutations associated with the transmissible mutant are indicated by solid black dots and are labelled. **b**, The receptor-binding site of the transmissible-mutant HA containing human receptor with the three sugars coloured in red (GlcNAc-3), blue (Gal-2) and yellow (Sia-1). The structural elements that form the edges of

the receptor binding site are labelled: 130-loop, 190-helix and 220-loop. **c**, The receptor-binding site of a second monomer (rotated about the vertical axis) showing bound avian receptor. **d**, The components of the human (left) and avian (right) receptor and their distinctive linkage types in the upper part and an overlap of the same two ligands shown in orientation of **c** in the lower part. **e–h**, The receptor-binding sites of wild-type VN1194 and tyTy H5 HA with bound human receptor (**e** and **g**, respectively) viewed in the same orientation of the complex as **b**, and wild-type VN1194 and tyTy H5 HA bound to avian receptor (**f** and **h**, respectively) in the same orientation of the complex as **c**.

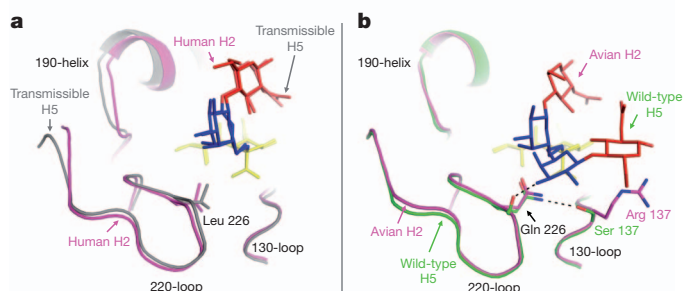


Figure 3 | Comparison of H5 and H2 HA complexes with human receptor. **a**, The receptor-binding site of transmissible-mutant HA (grey) overlapped with human H2 HA (purple) viewed in the same orientation as Fig. 2b showing the close similarity in the conformation of the human receptor in the two complexes. **b**, A similar view of the overlap of the receptor-binding complexes shows that the human receptor adopts a substantially different conformation in the wild-type H5 HA (green) than in the avian H2 HA (purple) complex.

with the wild-type H5 HA complexes. This mode of binding avian receptor by transmissible-mutant HA is shared by both H2 and H3 pandemic virus HAs^{4,5} (Supplementary Fig. 1).

In addition to the Gln226Leu mutation, transmissible-mutant HA contains three other amino acid substitutions: Asn158Asp, Asn224Lys and Thr318Ile (Fig. 2a and Supplementary Fig. 1). Asn158Asp results in the loss of a site for glycosylation in clade 1 H5 HAs (many clade 2 H5 HAs also lack this glycosylation sequon). Simple modelling studies based on our structure of the complexes formed between wild-type and the transmissible-mutant HA with human receptor suggest that the presence of a large carbohydrate side chain at residue 158 might sterically block receptor binding to cell-surface sialic acids¹⁸ (Supplementary Fig. 1a). Mutation of the glycosylation site would remove this potential impediment and thus facilitate the observed transmissibility in ferrets^{1,2}.

Residue 224 is located on the rim of the receptor-binding site (Fig. 2a, b and Supplementary Fig. 1) with its side chain oriented away from the site. It is not apparent that it influences receptor binding directly. However, the substitution of a lysine residue at this position introduces a basic residue that may, like similar mutations that have been described in H5 viruses isolated from humans¹⁹, enhance virus binding to cell surfaces through non-specific electrostatic interactions. It is notable in this regard that in two of the respiratory-droplet-transmissible-mutant H5 viruses that have been described, in addition

to the Gln226Leu amino acid substitution, there have been mutations that introduce basic amino acids: Gln196Arg³ and Asn224Lys¹. These observations suggest that increases in positive charge at the membrane-distal surface of HA may be favourable for non-specific interaction with negatively charged cell-surface components^{20,21}. Given the relatively modest specific binding that the transmissible-mutant HA displays towards human receptors, such a non-specific contribution may favour transmissibility.

Residue 318 is 70 Å away from the receptor-binding site (Fig. 2a and Supplementary Fig. 1) and is unlikely to influence receptor binding directly. However, it has been shown to increase the thermal stability of the transmissible-mutant HA¹, which is consistent with the known effects of mutations in similar locations in the intra- and inter-subunit interfaces of HA²². Our structure shows how the Thr318Ile mutation results in the isoleucine residue packing against the hydrophobic surface formed by Trp 21 in the 'fusion peptide' and Val 48 and Val 52 in the middle of helix A of the same subunit (Supplementary Fig. 1c) This interaction would stabilize the positions of both the fusion peptide and helix A, and as a consequence counteract the destabilizing effects of the three mutations in the membrane-distal locations²³, resulting in retention of the wild-type HA fusion pH¹.

As the first step in virus infection, HA receptor binding of appropriate affinity and specificity is an obvious requirement for the spread of infection between individuals and between species. For pandemic influenza viruses that originate in avian species the acquisition of both properties is required. Avian influenza viruses of the H5 subtype remain a threat because they have been shown to have the ability to infect humans and because the infections that they have caused in humans and in other species are extremely severe. Genetically, HAs of the H5 subtype are most closely related to HAs of the H1 and H2 subtypes, both of which have been the source of pandemic influenza virus HAs. However, the mechanism by which avian H1 viruses acquired the ability to infect and spread in humans is distinct from that used by avian H2 viruses. The H1 virus HA required the Glu190Asp and Gly225Asp mutations^{9,24,25} in the receptor-binding site. By contrast, the H2 HA required the Gln226Leu and Gly228Ser mutations^{9,14,26,27}. Importantly, in relation to observations on the transmissible-mutant virus, H2 precursor viruses, containing only the Gln226Leu mutation, were isolated in 1957 (ref. 9) and a similar H2 virus with the same single mutation has recently been shown to acquire transmissibility in ferrets²⁸. The structural consequences of the Gln226Leu substitution, widening of the gap between the 130- and 220-loops at the edges of the receptor-binding site, and generation of a hydrophobic environment at the base of the site, are very similar for the transmissible-mutant HA and for the HA of the H2 pandemic virus and are broadly similar to what occurs in the more distantly related HA of the H3 pandemic virus¹⁷ (Supplementary Fig. 1).

In some respects the properties of the transmissible mutant described here suggest that H5 viruses could take a similar evolutionary pathway in humans to that followed in 1957 and 1968 by avian H2 and H3 viruses. Thus, the preference of the transmissible mutant for binding to human versus avian receptors, and the structural manner by which it binds them, are highly characteristic of pandemic viruses. On the other hand, the transmissible mutant achieves its 200-fold preference for human receptor by essentially losing its binding potential for avian receptor. This property distinguishes it from known pandemic viruses. Whether or not these properties are shared by the other two transmissible H5 mutants^{2,3} remains to be determined.

The loss of affinity for avian receptor noted here also implies that the species in which a transmissible mutant of this sort may evolve would be restricted to either birds (such as quail²⁹) or mammals (such as pigs or humans³⁰) which could provide an abundance of human receptors. This restriction may also distinguish the transmissible mutant from viruses that were precursors of the H2 and H3 pandemics. Both avian H2 and H3 viruses with high avidity for both avian and human receptors

have been identified⁴, and they may have been directly transferred to humans from waterfowl, the main source of avian influenza viruses.

METHODS SUMMARY

Influenza viruses were grown in hens' eggs and purified by sucrose density gradient centrifugation, according to standard protocols. HA trimers were purified after proteolytic release from these viruses. HAs for wild-type A/Vietnam/1194/2004 (H5N1) (VN1194) and A/turkey/Turkey/1/2005 (H5N1) (tyTy) and the transmissible-mutant H5 were sub-cloned into a modified pAcGP67A vector that carries a tobacco etch virus (TEV) protease site, a trimerization foldon and a His tag. Protein was expressed in Sf9 cells and purified as described previously²¹. MST measurements were performed using a NanoTemper Monolith NT.115 instrument (NanoTemper Technologies GmbH). BHA samples (HA released from virus by bromelain digestion) were labelled with the amine-reactive dye NT-647 using the Monolith NT.115 Protein Labelling Kit RED-NHS. Binding curves were generated from the sodium salts of α 2,3-SLN and α 2,6-SLN (obtained from Dextra). For BLI, virus binding to defined receptor analogues was measured on an Octet RED biolayer interferometer (ForteBio). Biotinylated α 2,3-SLN and α 2,6-SLN were purchased from Lectinity Holdings Inc. Binding of viruses (at 100 pM) was measured at 25 °C in a 30–50-min association step. All solutions also contained 10 μ M oseltamivir carboxylate (Roche) and 10 μ M zanamivir (GSK) to prevent cleavage of the receptor analogues by the viral neuraminidase. The (relative) amount of virus bound to the biosensor at different RSLs was calculated from the amplitude of the response at the end of the association step. HA was crystallized according to standard procedures and HA–receptor complexes were prepared by soaking HA crystals in crystallization solution supplemented with 40 mM receptor analogues 2,3-SLN or 2,6-SLN. Diffraction data were collected at 100 K at the Diamond Synchrotron, processed with Mosflm or XDS and subsequent calculations carried out using the CCP4 suite. Structures were built with Coot and refined with Refmac.

Full Methods and any associated references are available in the online version of the paper.

Received 7 December 2012; accepted 5 April 2013.

Published online 24 April 2013.

1. Imai, M. *et al.* Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature* **486**, 420–428 (2012).
2. Herfst, S. *et al.* Airborne transmission of influenza A/H5N1 virus between ferrets. *Science* **336**, 1534–1541 (2012).
3. Chen, L. M. *et al.* *In vitro* evolution of H5N1 avian influenza virus toward human-type receptor specificity. *Virology* **422**, 105–113 (2012).
4. Liu, J. *et al.* Structures of receptor complexes formed by hemagglutinins from the Asian Influenza pandemic of 1957. *Proc. Natl Acad. Sci. USA* **106**, 17175–17180 (2009).
5. Eisen, M. B., Sabesan, S., Skehel, J. J. & Wiley, D. C. Binding of the influenza A virus to cell-surface receptors: structures of five hemagglutinin–sialyloligosaccharide complexes determined by X-ray crystallography. *Virology* **232**, 19–31 (1997).
6. Xu, R., McBride, R., Nycholat, C. M., Paulson, J. C. & Wilson, I. A. Structural characterization of the hemagglutinin receptor specificity from the 2009 H1N1 influenza pandemic. *J. Virol.* **86**, 982–990 (2012).
7. Hatta, M. *et al.* Growth of H5N1 influenza A viruses in the upper respiratory tracts of mice. *PLoS Pathog.* **3**, e133 (2007).
8. Connor, R. J., Kawaoka, Y., Webster, R. G. & Paulson, J. C. Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. *Virology* **205**, 17–23 (1994).
9. Matrosovich, M. *et al.* Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. *J. Virol.* **74**, 8502–8512 (2000).
10. Couceiro, J. N., Paulson, J. C. & Baum, L. G. Influenza virus strains selectively recognize sialyloligosaccharides on human respiratory epithelium; the role of the host cell in selection of hemagglutinin receptor specificity. *Virus Res.* **29**, 155–165 (1993).
11. Shinya, K. *et al.* Avian flu: influenza virus receptors in the human airway. *Nature* **440**, 435–436 (2006).
12. Calder, L. J., Wasilewski, S., Berriman, J. A. & Rosenthal, P. B. Structural organization of a filamentous influenza A virus. *Proc. Natl Acad. Sci. USA* **107**, 10685–10690 (2010).
13. Sauter, N. K. *et al.* Hemagglutinins from two influenza virus variants bind to sialic acid derivatives with millimolar dissociation constants: a 500-MHz proton nuclear magnetic resonance study. *Biochemistry* **28**, 8388–8396 (1989).
14. Gambaryan, A. S. *et al.* Specification of receptor-binding phenotypes of influenza virus isolates from different hosts using synthetic sialylglycopolymers: non-egg-adapted human H1 and H3 influenza A and influenza B viruses share a common high binding affinity for 6'-sialyl(N-acetyl)lactosamine. *Virology* **232**, 345–350 (1997).
15. Matrosovich, M. & Klenk, H. D. Natural and synthetic sialic acid-containing inhibitors of influenza virus receptor binding. *Rev. Med. Virol.* **13**, 85–97 (2003).

16. Ha, Y., Stevens, D. J., Skehel, J. J. & Wiley, D. C. X-ray structures of H5 avian and H9 swine influenza virus hemagglutinins bound to avian and human receptor analogs. *Proc. Natl Acad. Sci. USA* **98**, 11181–11186 (2001).
17. Ha, Y., Stevens, D. J., Skehel, J. J. & Wiley, D. C. X-ray structure of the hemagglutinin of a potential H3 avian progenitor of the 1968 Hong Kong pandemic influenza virus. *Virology* **309**, 209–218 (2003).
18. Wang, W. *et al.* Glycosylation at 158N of the hemagglutinin protein and receptor binding specificity synergistically affect the antigenicity and immunogenicity of a live attenuated H5N1 A/Vietnam/1203/2004 vaccine virus in ferrets. *J. Virol.* **84**, 6570–6577 (2010).
19. Yamada, S. *et al.* Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. *Nature* **444**, 378–382 (2006).
20. Arinaminpathy, N. & Grenfell, B. Dynamics of glycoprotein charge in the evolutionary history of human influenza. *PLoS ONE* **5**, e15674 (2010).
21. Lin, Y. P. *et al.* Evolution of the receptor binding properties of the influenza A(H3N2) haemagglutinin. *Proc. Natl Acad. Sci. USA* **109**, 21474–21479 (2012).
22. Ruigrok, R. W. *et al.* Conformational changes in the hemagglutinin of influenza virus which accompany heat-induced fusion of virus with liposomes. *Virology* **155**, 484–497 (1986).
23. Steinhauer, D. A. *et al.* Studies using double mutants of the conformational transitions in influenza hemagglutinin required for its membrane fusion activity. *Proc. Natl Acad. Sci. USA* **93**, 12873–12878 (1996).
24. Gamblin, S. J. *et al.* The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science* **303**, 1838–1842 (2004).
25. Stevens, J. *et al.* Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. *J. Mol. Biol.* **355**, 1143–1155 (2006).
26. Rogers, G. N. *et al.* Host-mediated selection of influenza virus receptor variants. Sialic acid- α 2,6Gal-specific clones of A/duck/Ukraine/1/63 revert to sialic acid- α 2,3Gal-specific wild type *in ovo*. *J. Biol. Chem.* **260**, 7362–7367 (1985).
27. Rogers, G. N. *et al.* Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. *Nature* **304**, 76–78 (1983).
28. Pappas, C. *et al.* Receptor specificity and transmission of H2N2 subtype viruses isolated from the pandemic of 1957. *PLoS ONE* **5**, e11158 (2010).
29. Gambaryan, A. S. *et al.* Receptor-binding profiles of H7 subtype influenza viruses in different host species. *J. Virol.* **86**, 4370–4379 (2012).
30. Baigent, S. J. & McCauley, J. W. Influenza type A in humans, mammals and birds: determinants of virus virulence, host-range and interspecies transmission. *BioEssays* **25**, 657–671 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to staff at the Diamond Light Source Synchrotron for assistance and beamline access under proposal 7707, E. Christodoulou and S. Vachieri for discussions on protein expression, the staff of the NIMR Large Scale Laboratory, L. Haire for assistance with crystallization experiments and S. Smerdon and P. Rosenthal for discussions. H.X. was supported by BBSRC (award number BB/E010806). This work was funded by the Medical Research Council through programmes U117584222, U117512723 and U117570592.

Author Contributions X.X., P.J.C., S.R.M., J.L., H.X., J.W.M., K.L., P.A.W., P.C., Y.K., J.J.S. and S.J.G. all performed experiments and contributed to the writing of the manuscript.

Author Information Structural data have been deposited with the Protein Data Bank under accession codes 4BGW, 4BGX, 4BGY, 4BGZ, 4BH0, 4BH1, 4BH2, 4BH3 and 4BH4. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.J.G. (sgambli@nimr.mrc.ac.uk) or J.J.S. (skeheljj@nimr.mrc.ac.uk).

METHODS

MST. MST measurements were performed using a NanoTemper Monolith NT.115 instrument (NanoTemper Technologies GmbH). HA samples were labelled with the amine-reactive dye NT-647 using the Monolith NT.115 Protein Labelling Kit RED-NHS. Labelling levels (generally in the range 0.3–0.4 dye molecules per HA monomer) were determined using $\epsilon_{280} = 74090$ (H3) and 83030 (H5) $\text{M}^{-1}\text{cm}^{-1}$ for the HA monomer concentration, and $\epsilon_{647} = 250,000$ $\text{M}^{-1}\text{cm}^{-1}$ for the dye concentration. Sodium salts of $\alpha 2,3$ -SLN and $\alpha 2,6$ -SLN (obtained from Dextra) were dissolved to a final concentration of 50–100 mM in PBS buffer containing 0.05% Tween-20 and labelled HA at a concentration of ~ 50 nM. This stock solution was then serially diluted 1:1 using the same buffer to give 12 working solutions with different sugar concentrations but the same fluorophore concentration. These solutions were then loaded into standard treated capillaries and MST measurements were made at 25 °C using 20% light-emitting diode power and 40% infrared-laser power. The laser-on time was 30 s and laser-off time 5 s. All measurements were made at least five times.

BLI. Virus binding to defined receptor analogues was measured on an Octet RED biolayer interferometer (ForteBio). Biotinylated $\alpha 2,3$ -SLN and $\alpha 2,6$ -SLN were purchased from Lectinity Holding. These were approximately 30-kDa polymers containing 20% mol sugar and 5% mol biotin linked to a polyacrylamide backbone. The polymers were immobilized on streptavidin biosensors (ForteBio) at concentrations ranging from 0.01–1.5 $\mu\text{g ml}^{-1}$. The RSL of the biosensor was calculated from the amplitude of the response at the end of the 5–10-min loading step. The maximum response at complete saturation was ~ 0.6 nm. Binding of viruses (at 100 pM) was measured at 25 °C in a 30–50-min association step. The buffer was 10 mM HEPES, pH 7.4, 150 mM NaCl, 3 mM EDTA and 0.005% Tween-20. All solutions also contained 10 μM oseltamivir carboxylate (Roche) and 10 μM zanamivir (GSK) to prevent cleavage of the receptor analogues by the viral neuraminidase. The (relative) amount of virus bound to the biosensor at different RSLs was calculated from the amplitude of the response at the end of the

association step. These measured amplitudes were normalized by dividing by the maximum response (typically 5–6 nm) and this normalized response was plotted as a function of the RSL (see Fig. 1a and Supplementary Fig. 2). These normalized virus binding response curves report the fractional saturation of the sensor surface (f) and smooth lines through the curves were generated by fitting the data to a simple variant of the Hill equation:

$$f = \frac{[\text{RSL}]^n}{[\text{RSL}_{0.5}]^n + [\text{RSL}]^n}$$

In which RSL is the relative sugar loading, $\text{RSL}_{0.5}$ is the relative sugar loading at half saturation ($f = 0.5$), and n is a Hill coefficient.

Crystallography. The HA of VN1194 was purified from vaccine strain virus RG14 grown in embryonated chicken eggs. HAs of tyTy and the transmissible mutant were expressed in Sf9 insect cells. Following previously published methods²² all H5 HAs were purified as trypsin released ectodomains in the final buffer (10 mM Tris-HCl, pH 8.0, 50 mM NaCl). VN1194 HA was crystallized from 0.1 M HEPES, pH 7.0, 0.05 M MgCl_2 , 28–30% PEG 550. tyTy HA was initially crystallized from Bis-tris propane, pH 7.5, 0.2 M K/NaPO₄, pH 7.0, 20% PEG 3350 and crystals were improved by seeding in Bis-tris propane, pH 7.5, 0.05–0.15 M K/NaPO₄, pH 7.0, 15–18% PEG 3350. The transmissible mutant was crystallized from 0.1 M HEPES, pH 7.0, 25–30% Jaffamine ED-2001. Ligand-soaking experiments were performed by soaking the crystals in crystallization solution supplemented with 40 mM receptor analogues for 16 h. VN1194 and transmissible-mutant crystals were frozen in liquid nitrogen directly from the drop. tyTy HA crystals were cryo-protected by addition of 25% ethylene glycol before freezing.

Diffraction data were collected on Diamond beamlines at 100 K, and processed in the XIA2 pipeline, before being scaled by Scala from the CCP4 suite. Structures were determined by molecular replacement in Phaser using the VN1194 structure as a search model (PDB code, 2IBX). MolProbity was used to validate the final structures.

CAREERS

TURNING POINT Teenager takes up venture capitalism in bid to help cure ageing **p.399**

@NATUREJOBS Follow us on Twitter for the latest news and features go.nature.com/e492gf

NATUREJOBS For the latest career listings and advice www.naturejobs.com

NEALE COUSLAND/SHUTTERSTOCK



Singapore has become a research hotbed, attracting scientists from around the world.

WORKING IN ASIA

The siren song of Singapore

The city state offers opportunities for intrepid scientists, but working there has drawbacks.

BY QUIRIN SCHIERMEIER

Jayne Thompson's truth is stranger than fiction. The young quantum-information theorist is studying the very foundations of existence. "Is there an objective reality out there?" she asks. "Or is reality strangely dependent on the actions of observers like us?"

The quest to understand the arcane laws that rule the world of quantum physics took Thompson from Australia to Singapore, the diminutive city state at the southern tip of the Malay Peninsula. Last August, she joined the 5-year-old Centre for Quantum Technologies at the National University of Singapore (NUS). There, 20 principal investigators and roughly 180 other researchers including post-docs and PhD students are probing the innermost workings of the Universe, unburdened by teaching obligations or excessive grant-writing duties. "It's a fantastic place, very open

and collaborative," says Thompson. "I am supported very generously and I can discuss my ideas with amazing people from different continents and cultures."

Singapore, a former British colony with an area of little more than 700 square kilometres, has become a hotbed of research. Lucrative funding opportunities, high salaries and a welcoming environment have lured leading researchers to the city state, and flocks of aspiring young scientists have followed. Institutions and organizations are seeking foreign talent, and about half of the 5,700-strong academic workforce at the NUS comes from overseas.

But newcomers will have to adjust to an insular scientific environment in which personal relationships can take on inflated importance — for better or worse. In tiny Singapore, it is hard to escape past interactions with funding agencies, ministries or academic administrators. And in a country where science is

governed by five-year plans, researchers must reckon with sometimes perplexing shifts in funding priorities.

KNOWLEDGE GROWTH

Singapore did not always have a booming knowledge economy. In the 1960s, when it became independent, the city state had its fair share of notoriety and social problems. Apart from a medical college founded by the British colonists, from which the NUS emerged, science was absent and public education was in its infancy. The nascent republic — made up mostly of indigenous Malays, and Chinese, Indian and Tamil immigrants — was not unified by language, history or religion. Singapore could not afford to be Singaporean: it had to be international to survive.

Science investments have been a major part of its strategy for economic growth. In the past 15 years, Singapore has become ►

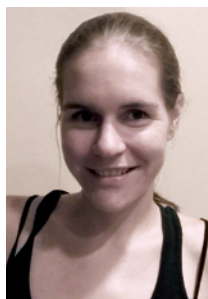
► a magnet for scientific talent. Its international ambience — superficially Western-seeming, except for the tropical climate — makes it easier for Europeans and Americans to live and work there than in China or Japan. “Accommodation is not hard to find and employment contracts are handled easily,” says Thompson. Her arrival “was a very smooth transition”.

All career levels are being recruited — from postgraduates to senior researchers (see ‘Welcome package’). The government’s Agency for Science, Technology and Research (A*STAR) offers scholarships and fellowships for graduate students and postdocs, and operates a programme under which PhD students from partner universities in Asia and the West spend up to two years at an A*STAR institute.

Postdocs are normally offered three-year contracts, with an option for another three-year term depending on a successful review. More-senior researchers negotiate a start-up package for the first one to two years, and then have to compete for grants.

Success rates for grant applications at A*STAR and other government funders vary between 15% and 25%, roughly the same as at the US National Institutes of Health. Generally, grants are about the same size as those in the United States and Europe — up to US\$500,000 per year. Salaries lie somewhere between those in Europe and those in North America.

International researchers tend to find Singapore welcoming. “I came with the ready-made notion of a country leaning a bit on the totalitarian side, where you would be hugely fined for jaywalking and the like,” says Artur



“I can discuss my ideas with amazing people from different continents and cultures.”

Jayne Thompson

York, notes continued restrictions on political opposition and freedom of speech — the city state has become more liberal in recent years.

PRODUCT PUSH

The Singaporean government has ruffled some feathers in the past three years by refocusing its science priorities — mainly in the life sciences.

In 2006, husband-and-wife cancer geneticists Neal Copeland and Nancy Jenkins left the US National Cancer Institute in Frederick, Maryland, for the A*STAR Institute of Molecular and Cell Biology (IMCB), which had made them a generous offer. “We both liked Asia and I was getting tired of working for the government,” says Copeland. The couple shipped their lab mice — and three

Ekert, the Polish-born director of the Centre for Quantum Technologies. “What I found is a relaxed and well-organized place — maybe a little boring, but safe, family-friendly and with a wonderful education system — where science is held in high esteem.” Although some authoritarian tendencies linger in Singapore — for example, Human Rights Watch, a non-governmental organization based in New

York, notes continued restrictions on political opposition and freedom of speech — the city state has become more liberal in recent years.

of their postdocs — to Singapore, where they hoped to stay until retirement. The scientific environment at the IMCB, which is located in the A*STAR-funded Biopolis research park, proved excellent, says Copeland. Students and co-workers were smart and hard-working, funding was almost unrestricted and the researchers had all the freedom they could have asked for. Within a year, Copeland was promoted to executive director of the institute and Jenkins to deputy director. The two felt much more engaged in scientific research than they had in the United States.

But things changed abruptly in September 2010, when the government decided as part of a new five-year research plan to increase commercialization of biomedical research and give funding priority to projects with industrial potential. The IMCB’s budget, which had been focused on basic research, was slashed by almost half. Copeland and Jenkins handed in their notice and returned to the United States, along with their postdocs.

Commercializing research was a longer-term goal when the couple first arrived in Singapore, says Copeland, who is now co-director of the cancer-biology programme at the Methodist Hospital Research Institute in Houston, Texas. “We could afford to quit when that changed because we’re approaching the end of our career. For many others it was more difficult.” Nevertheless, Copeland feels no resentment. “Singapore was by and large a wonderful experience,” he says. “You’ve got to be prepared that things can change overnight. But if you get a good job in science it is definitely worth going, especially if you’re young.”

For their postdocs Karen and Michael Mann, the Singapore experience was personally and professionally challenging. “Carving a life in a foreign country thousands of miles away from everything you know is emotionally taxing at times,” says Karen Mann. “At the time Jenkins and Copeland were getting ready to leave, there was a lot of uncertainty regarding funding, scientific leadership and scientific direction. We felt it was in our best interest to return to the United States.”

But it was excellent training for a career in science: the experience helped Karen and Michael Mann to forge valuable international collaborations and gain experience in designing and coordinating research projects. “We helped establish a mouse cancer-genetics lab from scratch as young postdocs, which is an invaluable experience for setting up our own labs,” says Karen Mann, who is still working under Copeland and Jenkins in Houston.

CULTURE CLASH

Some Western scientists complain that decisions by Singaporean funding agencies are not always transparent, and that the path to career advancement is unclear. Others have difficulties with the customs of Asian academics.

AWARDS SELECTION

Welcome package

Singaporean funding agencies offer a range of long-term support options for scientists from around the world. For example, the Singapore Translational Research Investigator Award from the National Medical Research Council is meant to attract experienced scientists to do translational and clinical research in Singapore. The five-year renewable grants include start-up costs, research support and salary (go.nature.com/jroafe).

Under the National Research Foundation Fellowship Scheme, postdoctoral and junior researchers from all nations and fields can apply for research grants of up to 3 million Singaporean dollars (US\$2.4 million) over five years. Recipients set up a group with at least one postdoctoral fellow and two PhD students to conduct independent research at a research organization of their choice (go.nature.com/xy6wwr).

The Agency for Science, Technology and

Research (A*STAR) Investigatorship aims to attract talented early-career researchers from all over the world. Awardees are appointed for three years, with provisions for a further three, and receive up to US\$500,000 a year (go.nature.com/ozl7os).

The Singapore International Graduate Award, a collaboration between A*STAR, Nanyang Technological University and the National University of Singapore, targets overseas students interested in doing a PhD in Singapore (go.nature.com/vxozwx).

Under the A*STAR Research Attachment Programme, PhD students from international partner universities can spend one or two years at an A*STAR institute under the joint supervision of staff from A*STAR and their home university (go.nature.com/ie1a3o).

The Singapore International Pre-Graduate Award funds overseas undergraduate students to spend 2–6 months at A*STAR institutes (go.nature.com/recdnj). **Q.S.**

Brendan Orner, a chemical biologist at King's College London, often felt alienated during his six years as an assistant professor at Nanyang Technological University in Singapore. "Students called me 'Sir' and wouldn't enter the room if I held the door for them," he says. "That was quite amusing, but the problems I had with superiors were less so." In one instance, he tried to get feedback on a failed grant application. But the university research-support officer thought that Orner was asking for the names of the reviewers. "I couldn't tell him that he had misunderstood me because that would have been an impossible affront," he says.

Confucian tradition, which is influential in Singapore, holds that senior people are not supposed to be challenged publicly. Westerners often find Asian restraint unsettling — whether in business meetings or in the seminar room. "You have to tell Asian students again and again that it is OK to challenge authorities," says Ekert. "But they are beginning to embrace Western concepts of discourse — and not only in science."

Barry Halliwell, the British-born deputy president for research and technology at the NUS, regularly has coffee or lunch with new recruits to get a sense of how they are getting on, and to help to settle any problems related to research administration, employment contracts or lab space. "Most settle in easily," he says. The odd problem — with lab space not being ready, for example, or with missing clearances for animal experiments — can normally be resolved in a short time.

Many foreign scientists will find a stay in Singapore a scientifically rewarding and character-building experience. "We have benefitted from the opportunity to work in a culturally diverse entity like A*STAR," says Karen Mann, who maintains productive collaborations that she started in Singapore with scientists in the United Kingdom, Australia and New Zealand.

Thompson is not yet sure where her inquiries into the nature of reality might lead her. "But Singapore," she says, "is nothing I'd wish to leave behind quickly." ■

Quirin Schiermeier is Nature's Germany correspondent.



"Things can change overnight, but if you get a good job in science, it is definitely worth going."

Neal Copeland

TURNING POINT

Laura Deming

Laura Deming was working in a research lab at the age of 12. Two years later, she was admitted to the Massachusetts Institute of Technology (MIT) in Cambridge and, at 16, she won one of the inaugural Thiel Fellowships, which encourage students to leave university and pursue business. As her fellowship winds down, Deming, now 19 and a partner at a venture-capital fund, remains devoted to pursuing anti-ageing therapies.

What prompted your interest in finding a cure for ageing?

When I was eight, my grandma had this wonderful spirit and wit, but she couldn't run around and play. I thought about all these people I know who have arthritic joints and disease, and whether there was anything that could be done to help them. If there was a way to make that happen, I wanted to work on it.

What was your first career turning point?

E-mailing molecular biologist Cynthia Kenyon at the University of California, San Francisco. She studies ageing, and I was reading about her, and thought, holy cow — someone is working on this stuff. So I contacted her and said, I'm 12, but I have to see if I could work in your lab. She was struck by my interest and let me come in as a volunteer. I experimented on genetically mutated strains of the roundworm *Caenorhabditis elegans*. Cynthia was my first mentor — she taught me how to think and be creative. She thinks as if there are no rules. Watching her changed how I am as a scientist in a very deep way. And working on this thing that I am passionate about changed my life.

What did you get from the Thiel Fellowship?

The fellowship — launched in 2010 by Peter Thiel, co-founder of PayPal — requires you to leave school and develop your business ideas with mentors in Silicon Valley, California. There is a lot of talk about human health in academic labs, but so many people in academia don't understand how the outside world works. Even graduate students don't have a picture of what happens once you make a discovery and how that could be developed into a drug. The fellowship was an extraordinary opportunity to see what happens from the venture perspective after the drug leaves the lab.

Will you return to MIT after your fellowship?

I'm a partner in the Longevity Fund in San Francisco, California, and I am going to stay in venture capital. My passion is to see that I'm getting therapies to market that will extend the



CHRISTOPHER RASCH

human lifespan. I could remain in academia and rely on a grant to support my research into those therapies, or I could see whether that research already exists, and do my best to help drugs to market.

Have you encountered age discrimination?

No. I have actually been surprised by the number of people I encounter who take me seriously. If you meet someone you might want to work with, the most important thing is that you're competent and have a good idea.

Should you take a lot of risks?

There is no downside to trying something. When I e-mailed Cynthia, the worst case was that she would read my e-mail and forget it, but the best case was that it would change my life.

What's the most effective way to network?

Find the best person in your field, think of a way that you can help them and get in touch. And don't overlook your friends. I worked with a graduate student at MIT who treated me like a true colleague, and I felt much more motivated as a result.

How can you get the most value from mentors?

Trekking out on your own is extraordinarily stupid. The best way to get ahead is to find the smart people and learn everything that you can from them. Apply those insights to what you're doing. You will avoid a lot of mistakes.

Do you have other advice for young scientists?

Be clear with yourself about what exactly you want to do in the long term. Know how you want to affect the world and what you are passionate about. Ignore boundaries and pursue what you want most to do. ■

INTERVIEW BY KAREN KAPLAN

THE PLAGUE

Lessons on life.

BY KEN LIU

I'm in the river fishing with Mother. The sun is about to set, and the fish are groggy. Easy pickings. The sky is bright crimson and so is Mother, the light shimmering on her shkin like someone smeared blood all over her.

That's when a big man tumbles into the water from a clump of reeds, dropping a long tube with glass on the end. Then I see he's not fat, like I thought at first, but wearing a thick suit with a glass bowl over his head.

Mother watches the man flop in the river like a fish. "Let's go, Marne."

But I don't. After another minute, he's not moving as much. He struggles to reach the tubes on his back.

"He can't breathe," I say.

"You can't help him," Mother says. "The air, the water, everything out here is poisonous to his kind."

I go over, crouch down, and look through the glass covering his face, which is naked. No shkin at all. He's from the Dome.

His hideous features are twisted with fright.

I reach over and untangle the tubes on his back.

I wish I hadn't lost my camera. The way the light from the bonfire dances against their shiny bodies cannot be captured with words. Their deformed limbs, their malnourished frames, their terrible disfigurement — all seem to disappear in a kind of nobility in the flickering shadows that makes my heart ache.

The girl who saved me offers me a bowl of food — fish, I think. Grateful, I accept.

I take out the field purification kit and sprinkle the nanobots over the food. These are designed to break down after they've outlived their purpose, nothing like the horrors that went out of control and made the world unlivable...

Fearing to give offence, I explain, "Spices."

Looking at her is like looking into a humanoid mirror. Instead of her face I see a distorted reflection of my own. It's hard to read an expression from the vague indentations and ridges in that smooth surface, but I think she's puzzled.

"*Modja saf-fu ota poiss-you*," she says, hissing and grunting. I don't hold the devolved phonemes and degenerate grammar against her — a diseased people scrabbling out an existence in the wilderness isn't exactly going to be composing poetry or thinking

philosophy. She's saying "Mother says the food here is poisonous to you."

"Spices make safe," I say.

As I squeeze the purified food into the feeding tube on the side of the helmet, her face ripples like a pond, and my reflection breaks into colourful patches.

She's grinning.

The others do not trust the man from the Dome as he skulks around the village enclosed in his suit.

"He says that the Dome dwellers are scared of us because they don't understand us. He wants to change that."

Mother laughs, sounding like water bubbling over rocks. Her shkin changes texture, breaking the reflected light into brittle, jagged rays.

The man is fascinated by the games I play: drawing lines over my belly, my thigh, my breasts with a stick as the shkin ripples and rises to follow. He writes down everything any one of us says.

He asks me if I know who my father is.

I think what a strange place the Dome must be.

"No," I tell him. "At the Quarter Festivals the men and women writhe together and the shkins direct the seed where they will."

He tells me he's sorry.

"What for?"

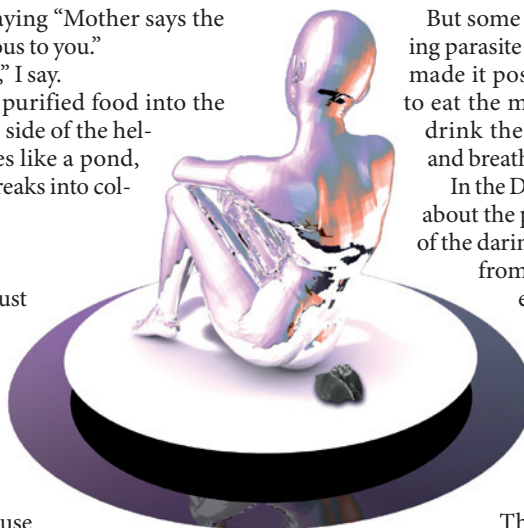
It's hard for me to really know what he's thinking because his naked face does not talk like shkin would.

"All this." He sweeps his arm around.

When the plague hit 50 years ago, the berserk nanobots and biohancers ate away people's skins, the soft surface of their gullets, the warm, moist membranes lining every orifice of their bodies.

Then the plague took the place of the lost flesh and covered people, inside and outside, like a lichen made of tiny robots and colonies of bacteria.

Those with money — my ancestors — holed up with weapons and built domes and watched the rest of the refugees die outside.



But some survived. The living parasite changed and even made it possible for its hosts to eat the mutated fruits and drink the poisonous water and breathe the toxic air.

In the Dome, jokes are told about the plagued, and a few of the daring trade with them from time to time. But

everyone seems content to see them as no longer human.

Some have claimed that the plagued are happy as they are.

That is nothing but

bigotry and an attempt to evade responsibility. An accident of

birth put me inside the Dome and her outside. It isn't her fault that she picks at her deformed skin instead of pondering philosophy; that she speaks with grunts and hisses instead of rhetoric and enunciation; that she does not understand family love but only an instinctual, animalistic yearning for affection.

We in the Dome must save her.

"You want to take away my shkin?" I ask.

"Yes, to find a cure, for you, your mother, all the plagued."

I know him well enough now to understand that he is sincere. It doesn't matter that the shkin is as much a part of me as my ears. He believes that flaying me, mutilating me, stripping me naked would be an improvement.

"We have a duty to help you."

He sees my happiness as misery, my thoughtfulness as depression, my wishes as delusion. It is funny how a man can see only what he wants to see. He wants to make me the same as him, because he thinks he's better.

Quicker than he can react, I pick up a rock and smash the glass bowl around his head. As he screams, I touch his face and watch the shkin writhe over my hands to cover him.

Mother is right. He has not come to learn, but I must teach him anyway. ■

Ken Liu is an author and translator of speculative fiction. For more about him and his work, visit <http://kenliu.name> or follow @kyliu99.

ON NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm